AN INVESTIGATION OF THE EFFECT OF SPACING OF PRACTICE ON THE PERFORMANCE–EFFICACY RELATIONSHIP

A Dissertation

by

ALOK BHUPATKAR

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2007

Major Subject: Psychology

AN INVESTIGATION OF THE EFFECT OF SPACING OF PRACTICE ON THE PERFORMANCE–EFFICACY RELATIONSHIP

A Dissertation

by

ALOK BHUPATKAR

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Winfred Arthur, Jr.
Committee Members,	Mindy E. Bergman
	Stephanie C. Payne
	Christopher O. L. H. Porter
Head of Department,	Les Morey

December 2007

Major Subject: Psychology

ABSTRACT

An Investigation of the Effect of Spacing of Practice on the

Performance–efficacy Relationship. (December 2007)

Alok Bhupatkar, B.A., University of Poona;

M.S., Emporia State University

Chair of Advisory Committee: Dr. Winfred Arthur, Jr.

The objective of the current study was to investigate the relationship between training performance and self–efficacy using a longitudinal design (approximately 11 weeks) in the context of massed and distributed practice. Limited attention in the training performance and efficacy literature has been paid to the spacing of practice (massed and distributed). However, it is conceivable that both the spacing of practice as well as the time frames over which it operates could influence the performance and efficacy relationship. Based on the practice schedule (massed versus distributed) and longitudinal study design, it was posited that the nature of the performance and efficacy relationship is likely to be quite different during two phases of learning (acquisition and reacquisition). Data were obtained from 198 undergraduate students over an 11–week training protocol using a 2 (distributed versus massed acquisition) × 2 (distributed versus massed reacquisition) × 16 (session) mixed design. Contrary to the first set of hypotheses, results indicated that the performance and efficacy relationship did not vary as a function of practice protocols (massed versus distributed) during acquisition and reacquisition. Also, no support was found for the hypothesis that the performance and efficacy relationship will vary as a function of whether the practice condition during acquisition is the same or different from the practice condition during reacquisition such that the relationships will be stronger when the practice condition is the same as opposed to when it is different. However, support was found for the hypothesis that when past performance is controlled the unique contribution of self–efficacy to subsequent task performance will be attenuated. Implications of these findings for research on the performance and efficacy relationship and training practice are discussed.

ACKNOWLEDGEMENTS

First, I wish to express my sincere gratitude to the chair of my committee, Winfred Arthur, Jr., for his guidance throughout this dissertation process. His expertise and moral support were greatly appreciated. He was a true mentor through this entire graduate school career. I would also like to thank Stephanie Payne, Mindy Bergman, and Christopher Porter for serving as members of my committee. Their expertise, suggestions, time, and support were also truly appreciated. In addition, I would also like to extend my special gratitude to Oiman Kwok who guided me through the data analysis process and taught me the fundamentals of hierarchical linear modeling. I also would like to thank Anton Villado who worked together with me in preparing the lab materials and all the undergraduate students who helped in the data collection process. Last, but not the least, I would like to thank my parents and Wendy Olson for their constant support and love.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The objective of the current study was to investigate the relationship between training performance and self–efficacy using a longitudinal design (approximately 11 weeks) in the context of massed and distributed practice. Limited attention in the training performance and efficacy literature has been paid to the spacing of practice protocols (massed and distributed). However, as discussed in subsequent sections of this dissertation, it is conceivable that both the spacing of practice, as well as the time frames over which it is studied, could influence the performance and efficacy relationship. An accepted doctrine in the spacing of practice literature is that distributed protocols are more effective than massed (Donovan & Radosevich, 1999). However, the effectiveness of distributed over massed practice is still unclear in the context of complex tasks and a longitudinal design that includes a nonuse period.

Furthermore, since Bandura's (1977) seminal self–efficacy work, the nature of the performance and efficacy relationship has splintered into two major schools of thought. The first, a school of thought founded by Bandura and others (Bandura & Wood, 1989; Locke & Latham, 1994; Stajkovic & Luthans, 1998) posits self–efficacy as the cause of performance. However, more recently another school of thought by Vancouver and others (e.g., Ackerman, Kanfer, & Goff, 1995; Arthur, Bell, & Edwards, 2007a; Heggestad & Kanfer, 2005; Judge, Jackson, Shaw, Scott, & Rich, 2007; Richard, Diefendorff, & Martin, 2006; Vancouver, Thompson, Tischner, & Putka, 2002;

This dissertation follows the style of *Journal of Applied Psychology.*

Vancouver, Thompson, & Williams, 2001) suggests that the best predictor of future performance is past performance and self–efficacy explains very little or no variance above past performance when examined longitudinally and actually shows a negative relationship to performance when examined within–individuals. This recent surge of research in performance–efficacy literature has provided a definitive answer to the performance and efficacy relationship when examined longitudinally within–individual.

However, what is less clear is the moderating effect spacing of practice protocols (distributed and massed) may have on the performance and efficacy relationship over time. To the best of my knowledge, this is one of the first studies that integrates the spacing of practice protocols with the performance and efficacy literature. In this process, the major objective of this dissertation was to contribute to both literatures and attempt to explain ambiguities in the self–regulation theories (e.g., social cognitive theory by Bandura, 1997) and Powers' (1973) control theory that have for decades investigated the performance and efficacy relationship. Some scholars (e.g., Vancouver et al., 2001) argue that the issue lies with the level of analysis (between–persons versus within–persons). Bandura and others (e.g., Stajkovic & Luthans, 1998) have predominantly used between–persons level of analysis and have found positive correlational relationships between performance and efficacy. Yet, Vancouver and colleagues (e.g., Heggestad & Kanfer, 2005) have argued that this positive relationship may be a function of the effect of performance on efficacy and not the other way around. In this study, both, between–persons correlational and within–persons across time analyses were addressed. Furthermore, because the performance and efficacy relationship in the current study was investigated across two phases of learning (acquisition and reacquisition) and over an 11–week long training protocol, the hypotheses were presented at the within–persons across time level of analysis.

Defining Self–Efficacy

Wood and Bandura (1989) define self–efficacy as “beliefs in one’s capabilities to mobilize the motivation, cognitive resources, and courses of action needed to meet given situational demands” (p. 408). Traditionally, the relationship between training performance and self–efficacy has been explained from the social cognitive theoretical standpoint (Bandura, 1997). Social cognitive theory posits a triadic reciprocal causal model in which behavior, cognitions, and the environment all influence each other in a dynamic manner (Bandura, 1986; 1997). The theory is based on the evaluative and agentic properties of human self–regulation that include, but are not limited to: (a) proactive adoption of aspirant standards, (b) self–appraisal of personal efficacy to fulfill particular goal challenges, (c) anticipatory regulation of the strategies and effort needed to convert standards into reality, (d) affective self–evaluative reactions to one’s performance, and (e) self–reflective metacognitive activity focused on the accuracy of one’s efficacy appraisals (Bandura, 1997; Bandura & Locke, 2003). Recent meta–analyses (Sadri & Robertson, 1993; Stajkovic & Luthans, 1998) have revealed a positive correlation between self–efficacy and performance. The self–efficacy literature suggests that high self–efficacy causes individuals to set higher goals, thus increasing their subsequent performance (Bandura, 1997; Bandura & Wood, 1989). Bandura and Wood (1989) found that participants who managed simulated organizations under a cognitive set that organizations are controllable maintained a strong sense of self–efficacy, set increasingly challenging goals, and exhibited effective analytical strategies, which further enhanced organizational performance. In addition, a tenet of goal setting is that self–efficacy increases goal commitment and hence, has a positive effect on subsequent performance (Locke & Latham, 1994).

However, the concept of self–efficacy has been widely debated (Corrigan, 1990; Hawkins, 1992; Lee, 1989) since its inception by Bandura in 1977. Some (e.g., Bandura, 1982; Earley & Lituchi, 1991) argue that self–efficacy is the *cause* of behavior; whereas others (e.g., Hawkins, 1992) are of the view that self–efficacy is merely a *predictor* of behavior. For example, Bandura's (1977) experiment on snake phobia showed that greater increments in snake phobic participants' self–efficacy as a result of treatment led to greater and positive changes in behavior. However, Hawkins (1992) argues that the behavior change as a result of treatment that Bandura calls "enactive mastery" is actually direct experience. Direct experience is also referred by Bandura as enactive mastery and is one of the four main sources of self–efficacy (Bandura, 1997). According to Bandura, enactive mastery refers to the successful recent experiences that individuals use in order to formulate their efficacy beliefs. Successful performance experiences raise individuals' efficacy beliefs, whereas, repeated performance failures lower them (Bandura, 1997).

Hawkins further argues with reference to the learning principles, that treatment should be the salient independent variable and not self–efficacy in Bandura's snake phobia experiment. Hence, as a rebuttal to Bandura's (1977) results, Hawkins suggests that *training* and not self–efficacy is the actual cause of behavioral change in snake phobic patients. Also, Borkovec (1978) criticizes Bandura's (1977) self–efficacy construct by arguing that since behavioral change can be attributed to the existing learning principles (reinforcement and punishment) without making reference to unobservable cognitions, self–efficacy should be better viewed as a *consequence* rather than a *cause* of behavioral change. In fact, Bandura agrees that there are reciprocal links between performance and self–efficacy acknowledging that behavior may determine self–efficacy. He states that "performance mastery, in turn, can boost perceived self–efficacy in a mutually enhancing process" (Bandura, 1982, p. 128). Hawkins (1992) is particularly of the view that the performance and self–efficacy relationship is not an insoluble chicken and egg problem. Self–efficacy should have an origin and undoubtedly previous behavior is extremely critical. Hence, I take a conceptual position in the current study that self–efficacy is a consequence of performance and not vice–versa.

In similar vein, recent empirical research (e.g., Ackerman et al., 1995; Arthur et al., 2007a; Heggestad & Kanfer, 2005; Judge et al., 2007; Richard et al., 2006; Vancouver et al., 2002; Vancouver et al., 2001) based on Powers' (1973) control theory paradigm and Guthrie's (1935) recency principle has taken the position that the positive relationship found between self–efficacy and performance is due more to the effect of past performance on self–efficacy rather than the effect of self–efficacy on subsequent performance. Heggestad and Kanfer (2005) showed that when past performance was unadjusted prior to entry into a raw past performance model, self–efficacy explained little or no variance in subsequent training performance; however, when past performance was statistically adjusted, self–efficacy explained variance in current performance over and beyond past performance. In the raw past performance model, the effects of past performance and self–efficacy are considered at a given point in time without considering any prior influence of either self–efficacy on performance or performance on self–efficacy. Studies (e.g., Shea & Howell, 2000) using the raw past performance model have found the effect of past performance to increase and the effect of self–efficacy to decrease over time. Judge et al.'s (2007) meta–analysis suggests that across all studies and moderator conditions in the meta–analysis, the incremental validity of self–efficacy on task and job performance was attenuated in the presence of specified individual difference variables (i.e., personality, general mental ability, experience). Specifically, Judge et al. found that although self–efficacy was moderately correlated with performance, the predictive validity of self–efficacy was dramatically reduced when the specified individual differences were taken into consideration. Further, Judge et al.'s meta–analytic review showed that self–efficacy predicted performance in jobs or tasks low in complexity but not those of medium or high complexity. Furthermore, self–efficacy predicted task performance but not job performance because efficacy was originally conceptualized and defined by Bandura (1997) as task specific and considered a better predictor of narrow performance compared to broad performance measures (Judge et al., 2007).

Massed versus Distributed Practice

The spacing of practice has been considered to be an important component of learning since the early 1900s when Ebbinghaus (1913) published one of his seminal pieces on memory (Hertenstein, 2001). Practice spacing was formally introduced in the field of education in the form of Jost's law (McGeoch, 1935). The law states that "if two associations are of equal strength, but of different age, a new repetition has a greater value than the older one" (McGeoch, 1935, p. 140). Spacing of practice is a critical factor in integrative learning methods (Bretz & Thompsett, 1992; Hertenstein, 2001) and can be classified as being either massed or distributed.

Massed practice schedules refer to the practice protocols in which the intertrial intervals are short; whereas in distributed practice schedules the intertrial intervals are longer (DeCecco, 1968; Donovan & Radosevich, 1999). Thus in practice, it would seem that the massed versus distributed distinction in temporal terms is relative and not absolute. Much of the research (e.g., Lee & Genovese, 1988) which has involved simple motor tasks has demonstrated that distributed practice conditions result in higher learning and task performance than massed practice conditions. Indeed, this effect is aptly reflected in Krug, Davis, and Glover's (1990) description of the spacing effect as "the phenomenon in which material encountered on two different occasions with a lapse of time between the encounters is remembered *better* than the material studied for an equal amount of time on one occasion" (p. 366). This common notion that distributed practice is more effective than massed may be due to the accepted doctrine in the learning literature which posits that tasks are learned better with protocols with longer intertrial intervals than those with shorter intervals. However, it is worth noting that the preponderance of this research has used simple tasks. Theoretical explanations for this effect are discussed in the subsequent section.

Theoretical underpinning: Spacing of practice protocol. There are five theories in the learning literature that can be used to explain why distributed practice protocols work better than massed practice protocols (Shebilske, Goettl, Corrington, & Day, 1999). These theories are presented in the milieu of spacing of practice schedules. The first theory—Hull's (1943) theory of reaction inhibition (also called drive theory of motivation) posits that prolonged activity on a task results in a negative drive (reaction inhibition) that eventually suppresses performance on the task. Hull speculates that the negative drive is caused by fatigue in massed practice protocols because the length of breaks between trials in massed protocols are shorter compared to distributed protocols. According to Hull this length of breaks between trials cause the negative drive to disperse over time. A break interval between two trials disperses inhibition resulting in increased performance after the break. Due to these break intervals, the negative drive may become nonexistent in distributed practice schedules over time compared to massed schedules. The theory was specifically established to explain the disadvantages of massed practice within a trial, but can also be generalized between trials to suggest that reactive inhibition disappears to a large extent during longer intertrial intervals (Shebilske et al., 1999).

A second theory, the metacognitive theory by Björk (1994) suggests that high performance during the massed practice causes a metacognitive error of overconfidence which results in individuals learning the material less effectively. Specifically, these metacognitive errors of overconfidence result in high short–term performance (e.g., skill acquisition), but poorer long–term performance (e.g., retention and transfer). Such metacognitive errors are less likely to occur in distributive than massed schedules. Björk defines learning as permanent changes in understanding, comprehension or competence that support long–term retention and transfer, and performance as the current speed or accuracy of access to the knowledge and skills that are the targets of training. Based on the conceptual distinction between performance and learning presented by Björk (1994), he argues that individuals are overconfident about their performance scores while being unaware to the amount of learning that actually takes place during skill acquisition. Björk speculates that this overconfidence occurs more in massed schedules as compared to distributed because individuals in massed schedules are more focused on their performance scores and less on the actual learning processes that are occurring. Furthermore, Björk explains that because individuals in massed schedules have little or no time to reflect on their learning processes, they draw a false picture of their actual state of learning which in turn leads to overconfidence. The structure of the distributed schedules gives individuals an opportunity to reflect on their learning processes that occur during trials, hence making more accurate judgments about their learning which is beneficial for long–term performance. According to Björk (1994), there are certain conditions in training that put individuals at risk of overestimating the degree to which skills and information are actually learned. One of these conditions is the structure of practice schedules, specifically, massed schedules that may yield better performance in the beginning of training, but lead to poor performance during the retention and transfer of training. Furthermore, Björk posits that this overestimation of skills may occur when the performance is measured either subjectively or objectively.

In recent years, other theories have addressed the effectiveness of distributed over massed practice in the context of the contextual interference effect (Immink & Wright, 1998). The contextual interference effect refers to a learning phenomenon where interference during practice is beneficial to skill learning such that higher levels of this effect result in poorer practice performance than lower levels, but result in higher retention and transfer performance (Magill & Hall, 1990). This effect was first demonstrated by Battig (1966) for primarily verbal tasks. Battig argued that when certain tasks must be learned and the tasks are themselves difficult and presented under conditions of high interference, the result is delayed retention that is as good as or better than easier tasks learned under noninterference conditions. Battig argued that high levels of interference typically led to poor performance during acquisition, but when transfer or retention trials were included, high levels of interference led to *delayed retention* produced by high interference acquisition situation. Battig stated three sources of interference that could enhance learning: (a) the task, (b) the practice schedule, and (c) the task information processing engaged by the learner. Specifically, Battig argued that task similarity increased contextual interference and therefore increasing processing activity demands of the task. Battig was primarily referring to the similarity between verbal tasks. Contextual interference has also been introduced in experimental situations by manipulating the characteristics of practice schedules (e.g., time lag, number of intertrial intervals). Furthermore, the extent to which the learner engages in information processing will determine the contextual interference effect.

There was a considerable lull in the contextual interference research for several years until Shea and Morgan (1979) used motor tasks to study the contextual interference effect again. In one of Shea and Morgan's first experiments, the participants were required to learn to move their arm as quickly as possible through three different three–segment patterns that included responding to a stimulus light, picking up a tennis ball and knocking over three freely moveable wooden barriers, and lastly returning the ball to a final location. The retention and transfer test results supported the contextual interference effect as participants in the distributed schedule outperformed those in the massed schedule.

Although contextual interference effect has been primarily studied using verbal and motor tasks, an attempt is made in the current study to apply the theoretical principles of contextual interference effect to complex nonmotor tasks. To explain the contextual interference effect several hypotheses have been presented. The most prominent ones are the elaboration and reconstruction hypotheses posited by Shea and colleagues (e.g., Shea & Morgan, 1979; Shea & Zimny, 1983). Hence, the third theoretical position, which is related to the contextual interference effect is the *elaboration hypothesis*. The elaboration hypothesis consists of two distinct cognitive processing modes: (a) intratrial, and (b) intertrial. Evidence (e.g., Shea & Zimny, 1983) suggests that participants in the massed schedule are more prone to using intratrial processing; whereas, distributed schedule participants are more adept at using intertrial processing. Magill and Hall (1990) refer to processing as learning of the task information that occurs either between trials or within a trial. Intertrial processing is more effective than intratrial processing because it involves relational or associative processing between two or adjacent trials that enables the participant to integrate new task information learned in the new trial with the existing knowledge from the previous trial. Furthermore, according to the elaboration hypothesis, distributed schedules lead to more distinctive and elaborate memory representations than does massed practice because individuals use multiple and variable task information processing strategies (Wulf & Shea, 2002). Shea and Morgan posited that because different tasks that need to be learned exist together in the working memory, the structure of distributed conditions provides an opportunity for participants to compare different tasks which lead to more distinctive and elaborate memory representations than massed practice. Although there is no consensus on a clear definition of working memory in cognitive psychology literature, Baddeley (1986) defined working memory as a mechanism or system underlying the maintenance and processing of task–relevant information during the performance of a complex task. Working memory makes it possible for several pieces of information to be held in mind simultaneously and interrelatedly. Furthermore, Baddeley considers working memory to be a subcomponent of the overall memory system that allows temporary storage and manipulation of information required for complex tasks. However, unlike the overall memory system that includes short– and long–term memory mechanisms, working memory is limited in both the storage of task information and processing capacity (see Baddeley, 1986, for a detailed review on working memory).

The fourth theoretical principle in reference to the contextual interference effect is the *reconstruction hypothesis*. According to the reconstruction hypothesis, in contrast to the elaboration hypothesis, the contextual interference created by distributed schedules leads to forgetting of the action plan or strategy during skill acquisition. Action plan refers to the continuous evaluation of task information and task related strategies (Wulf & Shea, 2002). The distributed practice schedules then require the repeated reconstructions of action plans during skill reacquisition (retention and transfer)—something that is not necessary in the massed schedules as the action plans already exist in working memory. Due to the already existing action plans in the working memory for participants in massed schedules, massed schedules are initially more effective than distributed schedules, whereas, distributed schedules are more effective during the retention and transfer of skills. Hence, Wulf and Shea (2002) posit that the repeated reconstruction of action plans in distributed practice is responsible for learning advantages over time. Although the mechanisms in the elaboration and reconstruction hypotheses are different, they both posit the effectiveness of distributed schedules over massed.

Finally, the fifth theory that explains the superiority of distributed over massed practice is the component–process theory by Glenberg (1979). The component–process theory posits that shorter intertrial intervals (or less spacing between tasks) will result in fewer fluctuations within and between task components. According to encoding variability theory (Bower, 1972), less fluctuation is considered to be a major disadvantage because it results in poorer encoding. Glenberg explains this through component–process theory that assumes that a stimulus is represented by a multi–component episodic memory trace. Which components or features are included in the memory trace depends on the actual stimulus that is presented, the nature of the task, the individual's strategies, and the context in which the stimulus is presented.

Glenberg (1979) distinguishes between three types of components: (a) *contextual*—representing the context at presentation, (b) *structural*—representing relations and associations between tasks, and (c) *descriptive*—representing specific task features (see Glenberg, 1979, for a detailed review of the three components). Glenberg argues that the components differ to the degree to which they are included in memory traces representing different tasks and the probability that they vary between successive presentations of the same task. Furthermore, spacing of presentations is likely to lead to more contextual, structural, and descriptive components stored in the memory trace, hence, resulting in better encoding and performance. Because distributed practice schedules have longer intertrial intervals than massed schedules, distributed schedules are more likely to produce larger fluctuations (or variations) between task components which in turn, result in richer encoding. Related to this theory is an argument proposed by James (1890) that suggests the changes that take place during the longer training periods may eventually determine the amount of learning that has occurred. Learning is then characterized by the quantitative changes like task accuracy, speed, and amount of fatigue over long task trials and qualitative changes like the effort demanded by different tasks. James presented these arguments while providing an explanation of the process from novel to skilled performance in the context of habit development.

Summary of Theories

At this juncture, it is important to recognize that all the aforementioned theories have been extensively tested in the cognitive psychology literature using verbal and simple motor tasks only. To this end, it is also important to recognize the unique contribution of each theory and how it complements or contradicts other theories while also supporting the effectiveness of distributed schedules over massed. First, there are two theories that address the process in which the distributed practice structure disperses the negativity (e.g., fatigue) that arises as a result of prolonged task activity. Specifically, Hull's (1943) reactive inhibition theory suggested that the break intervals that are an integral part of distributed schedules result in the dispersion of negative drive over time and attenuation of fatigue caused by successive trials without breaks. This reactive inhibition theory complements Glenberg's (1979) and James (1890) one critical principle of components process theory. Specifically, James argued that the longer training trials increase the likelihood of fatigue while decreasing the arousal state of the individual that is required for consistently high performance over time. However, one major distinction between Hull's reactive inhibition theory and Glenberg's component–process theory is that Hull originally conceptualized reactive inhibition for tasks within a trial only, whereas, Glenberg's component–process theory was applied within and between trials to investigate the fluctuations within and between task components.

Furthermore, there are two theories that suggest massed practice schedules are more effective for short–term performance compared to distributed schedules, but not for long–term performance (i.e., retention and transfer). Specifically, Björk (1994) posited that there are certain conditions in training like massed schedules that put individuals at risk of overestimating the degree to which skills and information are actually learned. Because participants in the massed schedule are more focused on the performance outcome and may not have the opportunity to reflect on the learning processes (Björk & Björk, 1992), Björk argues that this lack of opportunity to reflect on the learning processes may lead to overestimation of skills. This overestimation according to Björk leads to errors of overconfidence which in turn results in high short–term performance, but poorer long–term performance. Complementing Björk's meta–cognitive theory is Wulf and Shea's (2002) reconstruction hypothesis that posits that repeated reconstruction of action plans is not needed in massed schedules during skill acquisition as the action plans already exist in working memory. As a result, participants in massed schedules will perform better than distributed participants initially, but as the training progresses individuals in distributed schedules will continue to reconstruct the action plans and will eventually outperform individuals from the massed schedule. This will lead to better retention and transfer performance.

In contrast to the reconstruction hypothesis, Shea and Morgan (1979) posited that because different tasks that need to be learned exist together in the working memory, the structure of distributed schedules provides an opportunity for participants to compare these different tasks which result in more elaborate memory representations. Shea and Morgan referred to this hypothesis as the elaboration hypothesis. Furthermore, based on the elaboration hypothesis, distributed schedules promote intertrial processing of task information which is more effective than intratrial processing, because it involves relational processing between two trials that enables the participant to integrate new task information learned in the new trial with the existing knowledge from the previous trial. This further increases the level of distinctiveness between variable task information for distributed schedule participants which results in better short– and long–term performance. Hence, according to elaboration hypothesis, distributed schedule participants are likely to produce better short– and long–term performance compared to massed schedule participants, whereas based on the reconstruction hypothesis, distributed schedule participants are likely to produce better long–term performance only. Furthermore, although, the two hypotheses present the effectiveness of distributed schedules over massed, the causal explanations for each of the hypotheses presented are very different.

Previous Research on Massed versus Distributed Practice

Furthermore, as previously noted, the conceptual difference between distributed and massed conditions is *relative* rather than *absolute*. Specifically, trainees in distributed practice protocols are exposed to a higher number of break periods between training sessions in comparison to those in the massed protocol (Arthur et al., 2005). For example, whether five 2–hour long sessions over five days during skill acquisition is best described as distributed or massed depends to some extent on what it is compared to. Thus, if compared to a single 10–hour long session, then the five 2–hour long sessions over five days are best described as “distributed.” However, if compared to ten 1–hour long sessions over ten days, then the five 2–hour long sessions over five days are best referred to as “massed” (Arthur et al., 2007b).

According to Dempster (1988) there are serious discontinuities in the literature on the spacing effect. First, one of the major problems in the literature is the use of different terminology to refer to similar but distinguishable phenomenon. For example, “lag effect” occurs when performance improves as a result of breaks between consecutive presentations. On the other hand, massed versus distributed practice effects refer to comparisons between spacings of zero (massed practice) and all spacings greater than zero (distributed practice). However, lag and massed–versus–distributed practice effects have been used interchangeably (Dempster, 1988). Dempster considers this interchangeable use of terminology to be a serious issue and one that needs consideration.

Spitzer (1939) empirically showed that if the interval between original learning and the first test in a series is too lengthy, then test spacing effects are likely to be vitiated. Specifically in one of his studies on retention, Spitzer presented eight groups comprised of approximately 400 middle–school children with articles to read. First, groups received Article A and were tested immediately on the same article (Test A). Second, the groups received Article B, but only groups one and two were given Test B immediately, whereas other groups received Test B at varying time intervals after the start of the experiment. The results of this experiment indicated that groups one and two scored significantly better on Test B compared to other groups who received Test B at varying time intervals. Spitzer concluded that immediate recall in the form of a test is an effective method of aiding the retention of learning.

Before Donovan and Radosevich’s (1999) meta–analytic review, the consensus was that distributed practice conditions resulted in better learning and performance than massed practice. However, this research was based almost exclusively on simple motor tasks such that when the tasks have been more complex in nature, such as school activities (Dempster, 1988) and the reading of textbook material (Austin, 1921), research has failed to show the superiority of distributed over massed practice. Specifically, Austin found that massed reading of text material (e.g., six times in one day) proved as effective as distributed readings (e.g., daily for six days) in tests of immediate recall.

Considering the past literature on the spacing of practice effects, Donovan and Radosevich's (1999) meta–analysis clearly delineated boundary conditions that influence the spacing of practice effects. Donovan and Radosevich address three major concerns with the practice spacing literature. Regarding the first area of concern, they state that the majority of the literature on practice spacing effects has focused on the learning and performance of simple motor tasks. Thus, very little is known about more complex cognitive tasks. Second, there is a lack of concern among researchers regarding the potential boundary conditions that might either facilitate or debilitate the effectiveness of practice spacing effects. Donovan and Radosevich addressed this concern in their meta–analytic review and concluded that the type of task (simple or complex), the length of the intertrial time interval, and the interaction between these two factors play a significant role in determining the magnitude of the spacing of practice effects.

The third area of concern relates to the conceptualization of training performance. In most studies training performance is conceptualized and measured as performance immediately following the end of the practice sessions (i.e., acquisition performance).

Very few studies have examined training performance after a considerable period of time has elapsed after the completion of the practice (i.e., retention performance).

Donovan and Radosevich (1999) investigated the effect of task type and operationalized it in terms of three components of the task being performed: overall task complexity, mental requirements, and physical requirements. First, overall task complexity was defined as the extent to which the task requires a number of distinct behaviors, the number of choices required in order to perform the task, and the degree of uncertainty involved in the performance of the task. Donovan and Radosevich defined mental requirements of the task as “the extent to which the task requires an individual to use or demonstrate mental or cognitive skills and abilities in order to be able to perform the task” (p. 798). Third, physical requirements were defined as “the degree to which the task requires the subject to use or demonstrate physical skills and abilities in order to perform or complete the task” (Donovan & Radosevich, 1999, p. 798).

Donovan and Radosevich (1999) performed a cluster analysis based on the three task components to arrive at a four cluster solution. Cluster 1 included tasks high in physical requirements, and low in mental requirements and overall task complexity. Tasks like rotary pursuit, ladder climbing, and typing fall into this cluster. Cluster 2 included tasks low in physical requirements, high in mental requirements, and average in overall task complexity. Tasks in this cluster include learning a foreign language, maze learning, and voice recognition. Cluster 3 included tasks with high physical requirements, low mental requirements, and high overall complexity. Gymnastic skills and balancing tasks are representative of this cluster. Finally, Cluster 4 included tasks that are high in all three components. Tasks that are archetypal of this cluster include air traffic controller simulation, milk pasteurization simulation, and music memorization and recall. Donovan and Radosevich found that the overall effect size comparing distributed and massed practice for Cluster 1 (d = 0.97; tasks low in complexity and mental requirements and high in physical requirements) was higher than the overall effect size found for Cluster 4 (d = 0.07; tasks high in complexity, mental and physical requirements). Thus, the overall complexity of the task was a critical factor in determining the overall superiority of distributed over massed practice.

In summary, Donovan and Radosevich (1999) found that the overall mean sample–weighted effect sizes for overall task complexity, mental and physical requirements of the task, methodological rigor of the studies, and intertial time interval for the studies was d = 0.46 (k = 112, N = 8980). With reference to the intertrial interval, Donovan and Radosevich found that as the intertrial interval between the distributed practice trials became shorter, the standardized mean differences between distributed and massed condition groups increased (0.16 to 0.71). According to Donovan and Radosevich, this unexpected result can be attributed to the fact that a large number of tasks examined in the meta–analytic review consisted of simple motor tasks and it seems that any additional time between the distributed practice trials could have been harmful to the subsequent performance. Furthermore, another plausible explanation may be that the time lags between distributed practice trials were not optimal enough to perform well on simple motor tasks. Time lags are discussed later under the longitudinal design section. Second, the results can also be attributed to Guthrie's (1935) principle of forgetting in which he argues that learning does not disappear because of a hiatus (or length of time) between two intervals but primarily because of the new learning that occurs during that period of time.

Role of Spacing of Practice Conditions in the Performance–Efficacy Relationship

According to Bandura (1997), one of the main sources of self–efficacy in social cognitive theory is enactive mastery experience. Enactive mastery experiences or direct successful experiences provide the most authentic evidence of whether one can draw together whatever it takes to be successful. Bandura further posits that failures undermine self–efficacy beliefs especially when failures occur before self–efficacy is established. On the other hand, successful performances build a strong belief in one's personal efficacy.

There are several conceptual bases for why the spacing of practice (massed versus distributed) can influence the performance and efficacy relationship. First, since massed practice schedules have short or no intertrial intervals, it is highly likely that individuals may not have enough time to reflect on their performance between trials and thus form their self–efficacy beliefs. In contrast, the distributed practice schedules which entail longer intertrial intervals provide individuals with more opportunity to reflect on their performance between trials and may also provide more opportunity to help develop their self–efficacy beliefs. Second, the strength of the performance and efficacy relationship may be explained by the recency principle in the context of the spacing of practice. Specifically, according to Guthrie (1935), the recency principle suggests that when individuals are confronted with a stimulating situation or experience that closely resembles an earlier one, the individuals are more likely to react as they did previously and most recently. The task–specific nature of self–efficacy (Bandura, 1997) can also be explained through the recency principle where self–efficacy is more likely to be related to the most recent performance than a more distal performance. Humphreys (1960) was the first to apply the term simplex pattern to the correlation matrices of longitudinal data. Based on the recency principle, Humphreys argued that a simplex pattern in a correlation matrix exists when the largest correlations occur between temporally adjacent performance scores, with the correlations decreasing in magnitude as the number of intervening performance periods increase. Relatedly, according to Bandura's (1997) enactive mastery, individuals build efficacy beliefs based on their recent successful past performances. Bandura's enactive mastery (one of the four main sources of self–efficacy) is based on the Guthrie's recency principle. Integrating spacing of practice literature with performance and efficacy relationship, one may empirically investigate a few questions. For example, does the strength of the performance and efficacy relationship vary as a function of practice spacing? One could reasonably posit that compared to distributed practice, this relationship will be weaker in massed practice schedules because under massed practice, individuals may not have enough time to reflect on their performance between two performance periods.

According to Guthrie (1935), forgetting is more likely a matter of what an individual does during the intertrial interval instead of the length of time between the two intervals. This argument is best presented by Guthrie (1935) when he comments that "it is not time that robs beauty of its charm, but preoccupation with other affairs in its presence" (p. 102). In other words, learning does not disappear as a result of a mere lapse of time. It is only when that lapse of time includes new learning that interferes with the old learning.

Performance and Efficacy Spirals

In massed practice schedules, individuals may have little or no time to develop their sense of efficacy and I posit that the performance and efficacy spirals will be more pronounced in the distributed condition compared to the massed. Lindsley, Brass, and Thomas (1995) posited that the performance–efficacy relationship is cyclical in nature such that performance affects efficacy which in turn affects performance and so on. Hence, Lindley et al. define performance–efficacy spirals as the cyclical nature of the performance and efficacy relationship over time. These iterative loops are also called *deviation amplifying* in which a deviation in one variable (e.g., lower performance) leads to a similar deviation in another variable (decrease in efficacy) which continues to amplify over time. The cyclical nature of the performance–efficacy relationship may result in either an upward spiral (increasing performance and efficacy) or a downward spiral (decreasing performance and efficacy). Positive or negative performance outcomes may influence an individual's self–efficacy, thus creating certain demands for subsequent performance. The result may be an upward or downward spiral that is consistent over at least three task trials (Lindsley et al., 1995). For example, in an upward spiral, high performance will lead to high self–efficacy which in turn will lead to high performance for at least three trial cycles. For downward spiral, low performance will lead to low self–efficacy which in turn will lead to low performance.

However, another possible spiral pattern in contrast to upward and downward spirals, a self–correcting or *deviation counteracting* cycle occurs when a decrease in performance and efficacy is followed by an increase in either performance or efficacy (or vice versa). That is, "an analysis of performance allows one to make adjustments in future efforts that reverse the previous decrease (or increase) in performance and self–efficacy" (Lindsley et al., 1995, p. 650). Specifically, to meet the requirements for a self–correcting cycle, both performance and efficacy need to increase, but only one (either performance or efficacy) should decrease. For example, if both performance and efficacy at Trial 3 increase, then it should be followed by a decrease in either performance or efficacy at Trial 4. Or, both performance and efficacy need to decrease, but either performance or efficacy should increase to meet the self–correcting pattern. The relationship may fluctuate upward or downward over relatively short periods of time.

Lindsley et al. (1995) originally conceptualized performance and efficacy spirals in the context of skill acquisition. Further, in the context of a longitudinal design that includes acquisition and reacquisition, the performance and efficacy spirals may take longer to develop in skill reacquisition due to an extended period of nonuse between acquisition and reacquisition. Repeated performance outcomes over a short period of time foster quicker and routine information processing specifically automatic information processing (Daft & Lengel, 1984). However, a time lapse (e.g., nonuse period) between acquisition and reacquisition may lead to significant forgetting of task information. Also, Guthrie (1935) speculates that learning is not decayed due to merely a lapse of time, but when that lapse of time includes new learning that interferes with the old learning. Furthermore, Lindsley et al. argue that for a performance–efficacy spiral to establish itself, it takes at least three trials. Hence, after a nonuse period of eight weeks, it is speculated here that performance–efficacy spiral will take longer to re–establish itself.

In distributed practice schedules individuals may have more time to develop their sense of efficacy by mastering the task and being successful over time in the tasks they perform (Bandura, 1997). The longer intertrial interval during the distributed schedule provides an ideal opportunity for individuals to reflect on their past performance and as a result develop beliefs about efficacy. Based on the arguments presented earlier, it was predicted that the performance and efficacy relationship will be stronger for the distributed practice schedule compared to the massed. This investigation of the effect of practice spacing (massed and distributed) on the performance and efficacy relationship is considered to be a major contribution to the extant performance–efficacy literature.

Task Complexity

Another contribution to the extant literature is in the form of the performance task I used—a complex nonmotor task. Although an individual develops expertise with a complex task after multiple trials, the structural complexity of the task will not change. Moreover, from the training perspective, Schneider (1985) suggests that high performance skills (or tasks) have three main characteristics: (a) trainees must invest a considerable amount of time (i.e., greater than 100 hours) and effort to acquire a high–performance level, (b) training that produces high–performance skills is likely to result in high failure rates (greater than 20%), and (c) there should be substantive qualitative differences between the performance of the novice and the expert.

Furthermore, tasks may be classified as complex or simple based on the mental and physical requirements needed to perform them (Donovan & Radosevich, 1999). The majority of studies investigating the performance and efficacy relationship have used simple learning tasks like noun–pair lookup tasks (Heggestad & Kanfer, 2005). Addressing this issue, Heggestad and Kanfer note that “perhaps self–efficacy would have emerged as an important predictor if learning had been more difficult and required more practice time to stabilize” (p. 91). The present study addresses these issues and empirically tests the performance and efficacy relationship in both distributed and massed schedules using a highly complex nonmotor task. The task for this study, Jane’s Fleet Command (Sonalysts, Inc., 2004, Air Force Research Laboratory ver. 1.55) is a complex nonmotor task. It is a PC–based real–time micro–simulation of modern naval warfare featuring ships, aircraft, submarines, and airbases on land. Fleet Command meets the requisite criteria for cognitive complexity and information processing demands (Ericsson & Charness, 1994; Schneider, 1985) that include short– and long–term memory load, high workload, dynamic attention allocation, decision making, prioritization, and resource management.

One of the major purposes of the spacing of practice schedules has been to integrate learning (i.e., skill acquisition) and retention (opposite of skill loss) of complex and cognitively demanding tasks (Arthur, Day, Bennett, McNelly, & Jordan, 1997). One of the limitations of the skill decay literature and training literature in general is the tendency to treat skill acquisition, decay, transfer, and reacquisition as separate phenomena which are subsequently studied independently (Arthur & Bennett, 1996; Schmidt & Björk, 1992). Schmidt and Björk have criticized the educational and training research for treating learning and retention as two independent phenomena. Schmidt and Björk have shown that manipulations that maximize skill acquisition may not necessarily lead to higher retention and transfer. In other words, protocols that maximize skill acquisition may not lead to best retention and transfer compared to other protocols that may have slower speeds of skill acquisition. Hence, these scholars have argued that acquisition and retention are actually inseparable and should be studied together.

Role of Time in Performance and Efficacy Relationship

Longitudinal design. The present study's final contribution to the performance and efficacy literature pertains to the use of a longitudinal study design. Specifically, the effect of practice spacing (massed versus distributed) on the performance and efficacy relationship was investigated using a longitudinal training protocol that consisted of one or two weeks of acquisition, an 8–week nonpractice interval, and half or one week of reacquisition. Here, it was empirically examined whether the effects obtained in the acquisition phase were similar to or different from those in the reacquisition phase. This is an important contribution to the literature, because it permitted an examination of the pattern of the performance and efficacy relationship over a considerable long period of time. At this juncture, many questions were empirically explored. For example, how did the spacing of practice protocols influence the performance and efficacy relationship over time? When past performance was considered, did the unique contribution of self–

efficacy to subsequent task performance decrease over time? With the intent to empirically investigate these questions, I considered the longitudinal design of the study to be a significant contribution to the performance–efficacy literature.

Many scholars (Ancona, Okhuysen, & Perlow, 2001; George & Jones, 2000; Mitchell & James, 2001) have acknowledged the importance of time in organizational research, specifically between predictors and performance criteria. Time may influence the performance and efficacy relationships for the two practice conditions (distributed versus massed) and also influence these practice conditions differently. Mitchell and James suggest that most theory involves simple relationships between X and Y (e.g., performance and efficacy) representing substantive variables other than time. Within this paradigm, there are certain conditions under which the theory informs methodology with respect to time. One of the main conditions that Mitchell and James propose that is noteworthy to the current study is the time *lag* between X and Y. Mitchell and James also speculated that an optimal time lag between two variables is dependent on task complexity, the structure and spacing of practice presentations, and the rate of change between the two variables X and Y during the time lag and over time. Hence, there is no one optimal time lag that can be used across all conditions. Furthermore, the time lag between self–efficacy administrations at time 3 and 4 was an 8–week nonpractice interval. To the best of my knowledge, there is no empirical evidence that indicates if and how long of a nonpractice interval negatively affects the performance and efficacy relationship over time.

In this study, the time lag between performance and efficacy measurement differed for the two practice conditions. For the distributed condition the time lag was larger than that for the massed condition. For example, the lag between performance at Session 2 and self–efficacy at Time 2 was 48 hours for the distributed condition as compared to 24 hours for the massed. The time lag for participants in the distributed condition provided more time to reflect on their performance, hence, it was hypothesized here that the performance and efficacy relationship will be stronger when the time lag is larger as compared to when the time lag is smaller.

Second, Mitchell and James (2001) posit that X and Y may change over time and it is important to document the rate of change. In the current study, the rate of change between performance and efficacy was examined using within–person multi–level analysis in which the sessions served as level one and participants served as level two in the multi–level analysis. Here, it was examined whether the rate of change in the distributed condition is similar or different from that in the massed condition for each participant. These analyses are presented in the Results section. Furthermore, Mitchell and James suggest that it is also important to examine the dynamic relationships between X and Y, in which X and Y both change. The dynamic change between performance and efficacy was examined by analyzing the unique contribution of each variable to subsequent performance across five time periods.

Finally, George and Jones (2000) presented the role of six time dimensions in theory building about people, groups, and organizations: (a) The past, future, and present and the subjective experience of time; (b) time aggregations; (c) duration of steady states and rates of change; (d) incremental versus discontinuous change; (e) frequency, rhythm, and cycles; and (f) spirals and intensity. These authors claimed that the role of time must be explicitly incorporated into the theory building process and not just treated as a boundary condition. George and Jones argued that time is an intrinsic property of consciousness and stated that "the stream of consciousness is ordered temporally, and all conscious and motivated information processing takes place within the flow of time" (p. 659).

Based on the aforementioned theories addressing the effectiveness of distributed schedules over massed, the extant literature presented on spacing of practice effects (Donovan & Radosevich, 1999), and the performance and efficacy relationship (Arthur et al., 2007a; Heggestad & Kanfer, 2005; Vancouver & Kendall, 2006) in the sections above, the following is hypothesized:

Hypothesis 1: The performance and efficacy relationship will vary as a function of the spacing of practice protocols (i.e., *distributed* and *massed*) such that (a) the relationship in the distributed protocol during *acquisition* will be significantly stronger than that for the massed protocol during *acquisition* and (b) the relationship in the distributed protocol during *reacquisition* will be significantly stronger than that for the massed protocol during *reacquisition*.

Based on the component–process theory, Glenberg (1979) posits that a stimulus is represented by a multi–component episodic memory and which components or features are included in the memory depends among many others (i.e., structural and descriptive), the *context* in which the stimulus was presented. According to Glenberg, spacing of presentations is likely to lead to more contextual, structural, and descriptive components being stored in the memory, hence, resulting in better encoding and performance. Hence, I hypothesized here that more spacing of presentations (i.e., more likely in distributed protocols) combined with similar contextual components during acquisition and reacquisition will result in better encoding and hence, in stronger performance and efficacy relationships over time. To this end, the following was hypothesized:

Hypothesis 2: The performance and efficacy relationship will vary as a function of whether the practice condition during acquisition is the same or different from the practice condition during reacquisition such that the relationships are stronger when the practice condition is the same as opposed to when it is different.

According to Powers' (1973) control theory, one derives motivation from the comparison of current states with the desired states. Specifically, Powers argued that one can use self–efficacy beliefs to construct a perception of one's current state. When individual's efficacy about the current state is high, the person is likely to reach the desired state sooner by using lower resources than when one has lower levels of self–efficacy. Powers merely argued that the effect of self–efficacy on performance would not always be positive. Recently, several scholars (Arthur et al., 2007a; Heggestad & Kanfer, 2005; Judge et al., 2007; Vancouver et al., 2001) have demonstrated that the positive relationship found between self–efficacy and performance is due more to the effect of past performance on self–efficacy than the effect of self–efficacy on subsequent performance. Consistent with Powers' (1973) theory, the following was hypothesized:

> *Hypothesis 3.* When past performance is accounted for, the unique contribution of self–efficacy to subsequent task performance will decrease over time.

METHOD

Participants

The initial sample consisted of 236 paid volunteers recruited from university and campus communities at Texas A&M University and the University of Oklahoma. As a result of attrition, the final sample was 198. The attrition rate was approximately 13% of the initial sample and was roughly equal across conditions; 15 and 14 participants withdrew from the distributed and massed conditions respectively. Data from 9 participants were also removed because these participants did not follow training instructions. The mean age of the sample was 20.92 (SD = 3.63). One hundred eighty–four (93%) of the participants were college students and 153 (77%) were men. Trainees were paid $16.00 per hour for their participation (i.e., 17 hours of research participation). To motivate trainees to be engaged in the training sessions, they competed for three bonuses of $100, $60, and $40 which were awarded to the three trainees with the highest average task performance scores within their practice condition at the end of the study.

For the hypotheses (Hypotheses 1a and 1b) with the current sample size of 198, a test that regressed self–efficacy on performance and practice condition, postulating a small effect size (R^2 = .03) and an alpha of .05 resulted in a power of .60. Follow-up analyses, using the same boundary conditions indicated that a sample size of 325 would be needed to achieve a power level of .80. To place this in some context, the same test postulating a medium effect size (R^2 = .15) and an alpha of .05 would need a sample size of only 68 to achieve a power level of .80. Thus the study sample size of 198 provides sufficient power (power = .99) to detect a medium effect size but insufficient power (power = .60) to detect a weak effect.

Measures

Performance task—Jane's Fleet Command™. The performance task used was Jane's Fleet Command (Sonalysts, Inc., 2004, Air Force Research Laboratory ver. 1.55). Fleet Command is a PC–based real–time micro–simulation of modern naval warfare featuring ships, aircraft, submarines, and airbases on land. It provides the user with the ability to wargame carrier battle group strategy, tactics, and resource allocations and enables flexible, immersed, and interactive training. Fleet Command meets the requisite criteria for cognitive complexity and information processing demands (Ericsson & Charness, 1994; Ericsson, Krampe, & Tesch–Roemer, 1993; Schneider, 1985) including short– and long–term memory load, high workload, dynamic attention allocation, decision making, prioritization, and resource management. Consequently, it is an ecologically valid laboratory analogue of the types of cognitive, information processing, and decision making tasks and activities present in operational command–and–control environments in military, civilian first–responder, and other similar settings. This is highlighted by the use of Fleet Command for training purposes by several groups such as the U.S. Naval Academy, the Surface Warfare Development Group, the Surface Warfare Officers School Command, and the U.S. Naval Academy Division of Professional Development. Task performance was operationalized as—summed number of enemy aircraft and ships destroyed *minus* the summed number of friendly aircraft and ships lost.

Six different missions were used in the present study. The six missions were slightly modified to produce 3–4 variations of each mission depending on the number of times the mission was used across the different training sessions. Participants practiced using one variation of a mission and then tested using another variation of the same mission. The variations of a particular mission differed only in the location of own–side and enemy platforms. For example, a participant might have been attacked by an enemy fleet from the North during the practice mission, but attacked from the West during the test mission; otherwise, the missions were identical.

Self–efficacy. Participants' levels of efficacy for performing Jane's Fleet Command was assessed using a 6–item self–efficacy measure (see Arthur et al., 2006; Arthur et al., 2007a) that was developed following principles and guidelines recommended by Bandura (1997) for developing self–efficacy scales. Whetzel, Arthur, and Volz (in press) reported coefficient alphas of 0.94 and 0.95 for two administrations of this self–efficacy measure (see Appendix A for the self–efficacy scale). They also obtained a test–retest reliability of 0.95 for the self–efficacy scores.

Experimental Research Design and Procedure

The experimental design was a 2 (distributed versus massed acquisition) × 2 (distributed versus massed reacquisition) × 16 (session) mixed design. Acquisition and reacquisition conditions served as the between–subjects variables and session served as the within–subject variable. Although participants were randomly assigned to the practice conditions, due to difficulties encountered with accommodating participants' schedules, the distributed condition consisted of 102 participants and the massed consisted of 96. Table 1 provides an overview and summary of the research protocol. Upon being recruited to participate in the study, trainees were informed during a screening and scheduling session that they would be training to perform a complex decision making computer–based performance task. The screening session entailed the completion of a demographic and contact form, and a video/computer game experience measure. Although no one was eliminated on this basis, the intention was to exclude participants who reported extensive experience and familiarity with Fleet Command. Participants were selected into the study based on their availability and then randomly assigned to their specified practice condition.

The missions were presented to participants in an order of increasing difficulty. As the study progressed, the missions required participants to implement new strategies and systems, as well as those presented and learned in the preceding training sessions (e.g., SONAR, RADAR, specialized platforms, and/or specialized weapons). In addition, as the study progressed, mission goals required an increasingly greater degree of planning, the mission tasks required greater accuracy in navigation, and participants were required to coordinate and monitor more platforms (both own–side and enemy). The training strategy used here was based on the principles of *progressive–part training*. In progressive–part training, the trainees practice the first subtask in the first phase of training, then the first and second subtasks in the second phase; followed by the first, second, and third subtasks in the third phase, and so on (Wexley & Latham, 2002). Thus, the mission played during the baseline session, Session 1, and Session 2 was the least difficult (Mission 1), while the mission played during Session 11 was the most complex and difficult (Mission 6; also considered to be the transfer mission) and participants were required to implement all of the strategies and techniques presented in all of the preceding training sessions. Furthermore, whereas participants may have used proficiency in one area (e.g., sensor and weapons use, strike coordination, resource management, etc.) to compensate for a deficiency in another, effective performance on Mission 6 required proficiency in all tasks simultaneously.

Table 1

Overview of Training and Data Collection Procedures

ALL PARTICIPANTS	
Schedule Progression	Activity
Pre–training (Day 1)	Consent Forms Contact and Demographic Form Video/Computer Game Experience Measure
Pre–training (Day 2)	Assignment into conditions

Table 1 (continued)

DISTRIBUTED ACQUISITION		MASSED ACQUISITION	
Schedule Progression	**Activity**	**Schedule Progression**	**Activity**
Two Weeks of Skill Acquisition		**One Week of Skill Acquisition**	
Week 1		*Week 1*	
Tuesday	**Session 0** JFC Baseline **Session 1** JFC hands–on training JFC Practice and Test games (M1) Self–Efficacy (Time 1)	Monday	**Session 0** JFC Baseline **Session 1** JFC hands–on training JFC Practice and Test games (M1) Self–Efficacy (Time 1) **Session 2** JFC hands–on training JFC Practice and Test games (M2)
Wednesday	**Session 2** JFC hands–on training JFC Practice and Test games (M2)	Tuesday	**Sessions 3** JFC hands–on training JFC Practice and Test games (M2) **Sessions 4** JFC hands–on training JFC Practice and Test games (M2) Self–Efficacy (Time 2)
Thursday	**Session 3** JFC hands–on training JFC Practice and Test games (M2)		
Friday	**Session 4** JFC hands–on training JFC Practice and Test games (M2) Self–Efficacy (Time 2)		

Table 1 (continued)

DISTRIBUTED ACQUISITION		MASSED ACQUISITION	
Schedule Progression	**Activity**	**Schedule Progression**	**Activity**
Two Weeks of Skill Acquisition (9 sessions)		**One Week of Skill Acquisition (9 sessions)**	
Week 2		*Week 1*	
Monday	**Session 5** JFC Training Overview JFC Practice and Test Games (M3)	Wednesday	**Session 5** JFC Training Overview JFC Practice and Test Games (M3) **Session 6** JFC Practice and Test Games (M3)
Tuesday	**Session 6** JFC Practice and Test Games (M3)		
Wednesday	**Session 7** JFC Practice and Test Games (M4)	Thursday	**Session 7** JFC Practice and Test Games (M4) **Session 8** JFC Practice and Test Games (M4)
Thursday	**Session 8** JFC Practice and Test Games (M4)		
Friday	**Session 9** JFC Practice and Test Games (M5) Self–Efficacy (Time 3)	Friday	**Session 9** JFC Practice and Test Games (M5) Self–Efficacy (Time 3)

Table 1 (continued)

Distributed Reacquisition		**Massed Reacquisition**	
Schedule Progression	**Activity**	**Schedule Progression**	**Activity**
Five Days of Skill Reacquisition (7 sessions)		**Three Days of Skill Reacquisition (7 sessions)**	
Week 11		*Week 11*	
Monday	**Session 10** JFC Test Game (M5) **Session 11** JFC Test Game (M6) Self–Efficacy (Time 4)	Monday	**Session 10** JFC Test Game (M5) **Session 11** JFC Test Game (M6) Self–Efficacy (Time 4)
Tuesday	**Session 12** JFC Training Overview JFC Practice and Test Games (M2)	Tuesday	**Session 12** JFC Training Overview JFC Practice and Test Games (M2)
Wednesday	**Session 13** JFC Practice and Test Games (M3)		**Session 13** JFC Practice and Test Games (M3)
Thursday	**Session 14** JFC Practice and Test Games (M4)		**Session 14** JFC Practice and Test Games (M4)
Friday	**Session 15** JFC Practice and Test Games (M5) **Session 16** JFC Practice and Test Games (M6) Self–Efficacy (Time 5)	Wednesday	**Session 15** JFC Practice and Test Games (M5) **Session 16** JFC Practice and Test Games (M6) Self–Efficacy (Time 5)

Note: JFC = Jane's Fleet Command. M1–M6 = Mission 1–Mission 6. Missions were progressively more difficult and complex.

Prior to training on the first day of the study, participants completed the baseline mission of Fleet Command. Participants were given time to read the standard mission briefing, specifying the goals and objectives of the mission, followed by brief instructions on how to access the in–game help system in order to identify the keys and instructions necessary to operate Fleet Command. After completing the baseline mission, trainees then received instruction and tutorials on how to “play” Fleet Command. Training was delivered by an instructor who guided participants through the training sessions as participants followed along on their individual workstations using a fully functional Fleet Command mission. Following the training portion of the sessions, participants completed a practice mission followed by a test mission. Subsequent sessions followed the training sequence presented in Table 1. Sessions were scheduled to be an hour long and consisted of 20 minutes of practice and 25 minutes of testing—unless there was a tutorial or training which then utilized approximately 15 minutes and the practice and testing were then 15 minutes and 25 minutes long, respectively.

RESULTS

Descriptive Statistics

The means and standard deviations for the distributed and massed practice conditions for all the training sessions along with the associated effect sizes are presented in Table 2. The performance differences between massed and distributed practice conditions were statistically different for only Session 4 (d = .31). Descriptive information for self–efficacy scores across the five self–efficacy administrations is presented in Table 3. Figure 1 pictorially represents the performance scores across all sessions for massed and distributed practice conditions.

Performance–Efficacy Relationship Between Individuals

Table 4 provides the performance and efficacy correlations across all sessions for both distributed and massed conditions at the between–person level of analysis.

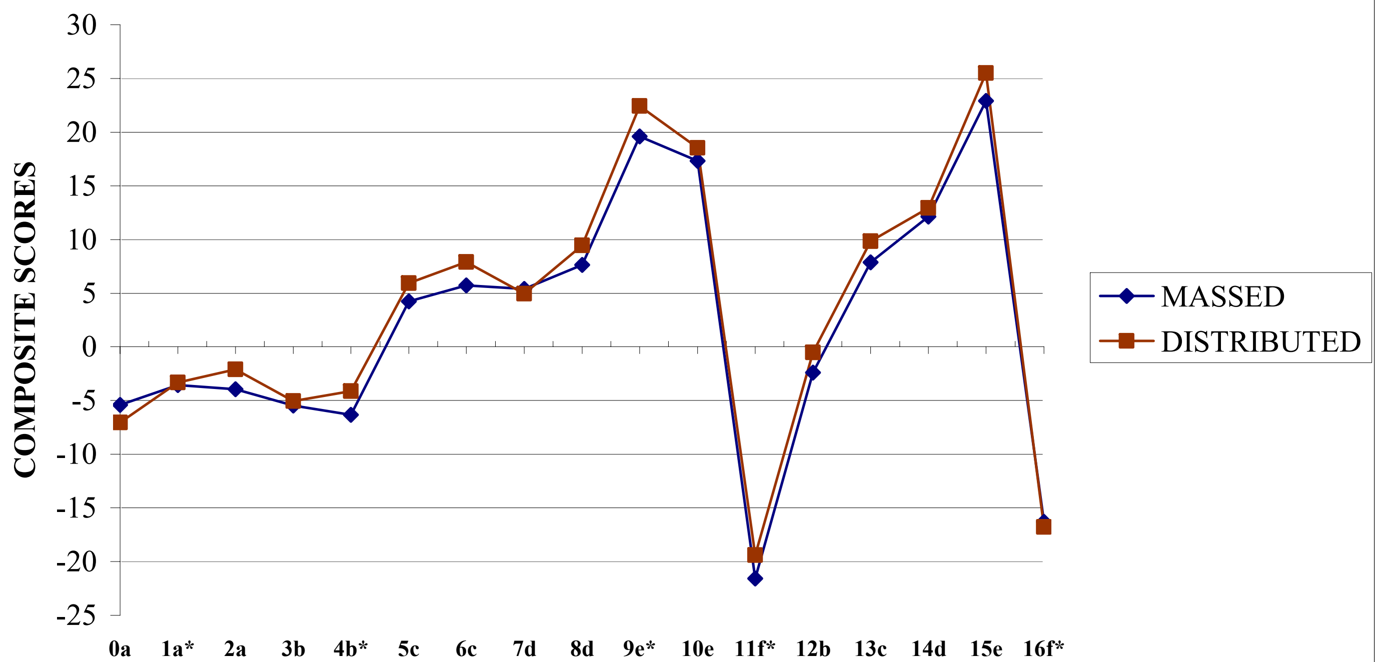
The results presented in Table 4 indicate that the performance and efficacy correlations for distributed condition were statistically stronger than the massed condition in only three (i.e., Sessions 4, 7, and 9) out of the 9 sessions during acquisition. A similar pattern of results was obtained for the reacquisition sessions as well, in which the performance and efficacy correlations were stronger for distributed than massed in only one (i.e., Session 14) out of the 7 sessions. The performance and efficacy correlations collapsed across distributed and massed conditions for all sessions are presented in Table 5. These results indicate that the performance–efficacy correlations were statistically significant for eleven out of 16 sessions. Figure 2 pictorially represents the findings regarding performance and efficacy correlations presented in Tables 4 and 5. These results provide descriptive information at the between–persons level of analysis only. In order to examine the first set of hypotheses, I used hierarchical linear modeling (HLM) at the within–person level of analysis that is presented next.

Table 2

Descriptive Statistics and Standardized Mean Differences for Composite Performance Scores for Acquisition and Reacquisition Schedules on all Fleet Command Sessions

Session	**Distributed**		**Massed**		***d***
	Mean	***SD***	***Mean***	***SD***	
Baseline					
0_A	-7.03	13.12	-5.41	9.46	-0.14
Acquisition					
>1_A	-3.31	12.84	-3.55	11.72	0.02
2_A	-2.09	8.63	-3.95	12.83	0.17
3_B	-5.05	5.46	-5.47	5.27	0.08
>4_B	-4.12	7.58	-6.32	6.55	0.31 *
5_C	5.93	9.19	4.24	9.45	0.18
6_C	7.91	8.54	5.73	9.06	0.25
7_D	4.97	12.60	5.41	13.71	-0.03
8_D	9.46	14.26	7.64	13.32	0.13
>9_E	22.45	13.46	19.61	14.48	0.20
Retention					
10_E	18.55	13.63	17.32	13.56	0.09
Transfer					
>11_F	-19.39	23.28	-21.59	26.19	0.09
Reacquisition					
12_B	-0.51	8.03	-2.39	8.03	0.23
13_C	9.85	6.64	7.87	8.32	0.26
14_D	12.93	12.38	12.12	15.82	0.06
15_E	25.50	12.33	22.91	14.40	0.19
Transfer					
>16_F	-16.75	19.17	-16.30	19.50	-0.02

Note. * $p < .05$. Subscript, uppercase letters indicate the mission performed in each session. In computing the *d*s, the distributed condition was treated as the experimental group ($n = 102$) and the massed condition as the control ($n = 96$). > in the first column indicates sessions during which self–efficacy was administered.



Note. *** indicates sessions during which self-efficacy was administered. Lowercase letters indicate mission performed in each session.**

Figure 1. Composite performance scores across all sessions for massed and distributed practice conditions.

Table 3

Descriptive Statistics for and Correlations Between Self-Efficacy Administrations across Massed and Distributed Conditions

Administration	*M*	*SD*	Time 1	Time 2	Time 3	Time 4	Time 5
Self–Efficacy Time 1 (M)	3.67	.71	*.91*				
Self–Efficacy Time 1 (D)	3.67	.62	*.90*				
Self–Efficacy Time 2 (M)	3.62	.81	.78	*.92*			
Self–Efficacy Time 2 (D)	3.72	.69	.72	*.89*			
Self–Efficacy Time 3 (M)	3.56	.78	.69	.72	*.90*		
Self–Efficacy Time 3 (D)	3.73	.73	.55	.65	*.91*		
Self–Efficacy Time 4 (M)	3.34	.85	.56	.61	.77	*.89*	
Self–Efficacy Time 4 (D)	3.49	.77	.43	.60	.71	*.90*	
Self–Efficacy Time 5 (M)	3.40	.84	.50	.57	.72	.86	*.90*
Self–Efficacy Time 5 (D)	3.55	.73	.45	.64	.76	.82	*.91*

Note: In parenthesis M = Massed, D = Distributed. Coefficient alpha reliabilities are located on the diagonal. All above correlations are significant at .01.

Table 4

Correlations Between Performance and Self–Efficacy across Sessions

	Performance (Jane's Fleet Command Session Number)																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Self-Efficacy																	
Time 1 (S1)																	
Massed	.06	.25*															
Distributed	.08	.01															
Time 2 (S4)																	
Massed			.27**	.08	.13												
Distributed			.12	.04	.28**												
Time 3 (S9)																	
Massed						.07	.20*	.22*	.30**	.30**							
Distributed						.14	.18	.34**	.27**	.41**							
Time 4 (S11)																	
Massed											.25**	.14					
Distributed											.25**	.15					
Time 5 (S16)																	
Massed													.16	.28**	.03	.23*	.07
Distributed													.16	.12	.28**	.14	.01

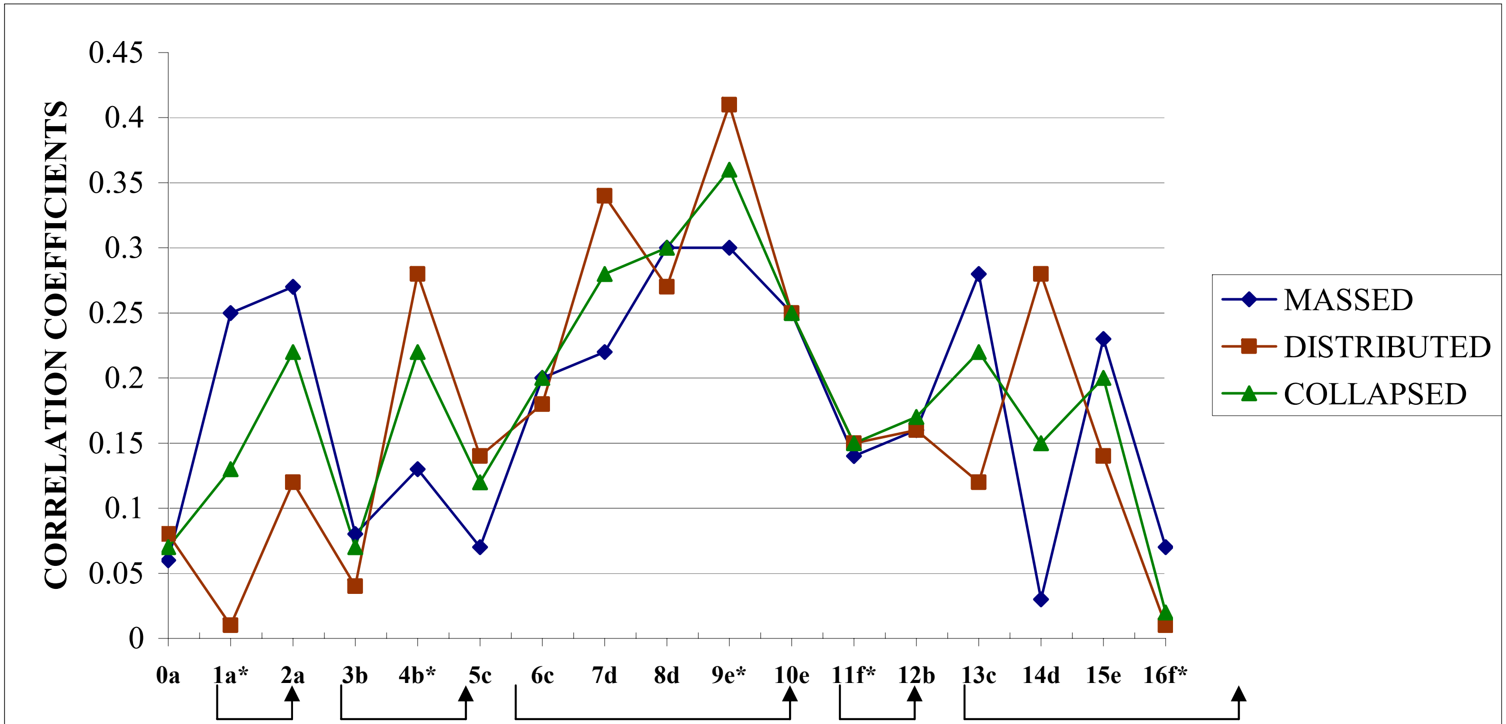
Note. S = Session. Time 1 through Time 5 refers to the administrations of the self–efficacy measure after task performance in Sessions 1, 4, 9, 11, and 16 respectively. * $p < .05$. ** $p < .01$.

Table 5

Correlations Between Performance and Self–Efficacy Collapsed across Conditions for all Sessions

	Performance (Jane's Fleet Command Session Number)																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Self-Efficacy																	
Time 1	.07	.13															
Time 2			.22**	.07	.22**												
Time 3						.12	.22**	.28**	.30**	.36**							
Time 4											.25**	.16*					
Time 5													.17*	.22**	.13	.20**	.02

Note. Time 1 through Time 5 refers to the administrations of the self–efficacy measure after task performance in Sessions 1, 4, 9, 11, and 16 respectively. * $p < .05$. ** $p < .01$



***Note.* * indicates sessions during which self-efficacy was administered. Lowercase letters indicate mission performed in each session.**

Figure 2. Performance and efficacy correlations across all sessions for distributed, massed, and collapsed across practice conditions.

Performance–Efficacy Relationship Within Individuals

The hypothesis (Hypotheses 1a and 1b) tests involved assessing the relationship between performance and self–efficacy scores for both distributed and massed practice conditions within a person across time. Hypothesis 1a posited that the performance–efficacy relationship will vary as a function of the spacing of practice protocol (i.e., *distributed* and *massed*) such that the relationship in the distributed protocol during *acquisition* will be significantly stronger than that for the massed protocol during *acquisition.* To examine this Hypothesis 1a, I used HLM (Bryk & Raudenbush, 1992). HLM assesses the covariation between variables, for each participant, across their three sessions (i.e., three administrations of self–efficacy during acquisition).

Because the hypothesis focused on the performance and self–efficacy relationship across sessions, session was the *first* level of analysis. Since individual differences are likely to confound the results if ignored, the performance and efficacy relationships assessed across sessions were examined for each individual, making individual the *second* level of analysis (Vancouver & Kendall, 2006). Hence, in this case, the individuals served as replicates. HLM then reports an average regression weight, gamma (γ), across individuals. Hypothesis 1a proposed that the interaction effect (i.e., performance × practice condition) would be significant across three sessions. Table 6 contains the average regression weight (i.e., γ) across all 198 participants. The γ coefficients for the interaction between performance and practice condition were not significant. Therefore, these results indicated that the performance and efficacy relationship did not vary as a function of practice protocols during acquisition and failed to provide any support for Hypothesis 1a.

Hypothesis 1b predicted that the relationship in the distributed protocol during *reacquisition* would be significantly stronger than that for the massed protocol during *reacquisition*. Similar to the analyses conducted for Hypothesis 1a, HLM procedure was used to test this hypothesis. In order to find support for Hypothesis 1b, the interaction effect (i.e., performance × practice condition) would have to be significant across two sessions during reacquisition. These results are presented in Table 6 below. Again, the γ for the interaction effect between performance and practice condition was not significant and the percentage of variance explained in self–efficacy was very small. Hence, the results presented in Table 6 failed to provide support for Hypothesis 1b.

Table 6

Hypotheses Tests at the Within–Person Level Analysis Using HLM

Variable	**γ**	***SE***	**% σ explained**
Self–efficacy as dependent variable: *Acquisition*			
Performance	.02	.03	1%
Practice Condition	.06	.09	6%
Performance × Practice Condition	.03	.04	2%
Self–efficacy as dependent variable: *Reacquisition*			
Performance	.04	.03	3%
Practice Condition	.01	.11	0%
Performance × Practice Condition	.03	.04	2%

Note: *N* = 198; Practice Condition was coded as Massed = 0, Distributed = 1. None of the above γ coefficients are statistically significant.

Hypothesis 2 predicted that the performance and efficacy relationship will vary as a function of whether the practice condition during acquisition is the same or different from the practice condition during reacquisition such that the relationships are stronger when the practice condition is the same as opposed to when it is different. Again, this hypothesis was examined by using HLM to compute γ in order to obtain an interaction effect between performance and four categories of practice condition (*category 1*: distributed during acquisition and distributed during reacquisition; *category 2*: distributed during acquisition and massed during reacquisition; *category 3*: massed during acquisition and massed during reacquisition; *category 4*: massed during acquisition and distributed during reacquisition). Because of the very nature of this hypothesis that required comparisons between practice conditions across the two phases of learning (i.e., acquisition and reacquisition), only reacquisition data were used while testing this hypothesis.

It was expected that the performance and practice category interaction would be significant across the two sessions during reacquisition. Results for this hypothesis are presented in Table 7. The γ for the interaction between performance and practice category was not significant ($\gamma = .01$, *ns*) and explained no variance in efficacy. These results indicate that the performance and efficacy relationship did not vary as a function of whether the practice condition during acquisition was the same or different from the practice condition during reacquisition. Hence, Hypothesis 2 was not supported.

Exploratory analysis. A within–person analysis using HLM was conducted to examine the effect of practice conditions (distributed versus massed) on performance across 16 performance sessions. In this HLM procedure for acquisition, performance session (i.e., 9 performance sessions) was used as first level and participants as second level of analysis. For reacquisition, seven performance sessions were used as first level of analysis and participants as the second level. Results of this analysis are presented in Table 8. These results indicate that the practice conditions explained 5% of variance during acquisition and 1% variance during reacquisition in performance across 16 performance sessions. However, the γ coefficients were not statistically significant for either phase of learning.

Table 7

Hypothesis Tests for Practice Category at the Within–Person Level Analysis Using HLM

Variable	**γ**	***SE***	**% σ explained**
Self–efficacy as dependent variable: *Reacquisition*			
Performance	.04	.05	2%
Practice Category	.01	.05	1%
Performance × Practice Category	.01	.02	1%

Note: *N* = 198; *Category 1*: distributed during acquisition and distributed during reacquisition; *Category 2*: distributed during acquisition and massed during reacquisition; *Category 3*: massed during acquisition and massed during reacquisition; *Category 4*: massed during acquisition and distributed during reacquisition. None of the above γ coefficients are statistically significant.

Table 8

Exploratory Within–Person Level Analysis Using HLM

Variable	**γ**	***SE***	**% σ explained**
Performance as dependent variable: *Acquisition*			
Practice Condition	.16	.09	5%
Performance as dependent variable: *Reacquisition*			
Practice Condition	.01	.11	1%

Note: $N = 198$; Practice Condition was coded as Massed = 0, Distributed = 1. None of the above γ coefficients are statistically significant.

Contribution of Self–Efficacy above Performance

Hypothesis 3 posited that when past performance is considered, the unique contribution of self–efficacy to subsequent task performance will decrease over time. To examine this hypothesis, in each regression, the current performance was regressed on past performance in Step 1 and self–efficacy in Step 2. It was expected that self–efficacy would explain little or no variance above prior performance across all four times. The results presented in Table 9 indicate that only at Time 2 did the self–efficacy ratings explain variance beyond prior performance and the change in R^2 ($\Delta R^2 = .03$) was statistically significant, $p < .01$. Except for criterion performance score at Session 2, past performance was a significant predictor of current performance in each of the remaining three regression analyses as R^2 ranged from .06 to .13 and were statistically significant. These results are consistent with Hypothesis 3 and indicate that when past performance is considered, the unique contribution of self–efficacy to subsequent task performance decreases. This finding is also consistent with a simplex pattern reflected in the intercorrelations between the performance scores reported in Table 10. A simplex pattern in a correlation matrix exists when the largest correlations occur between temporally adjacent performance scores, with the correlations decreasing in magnitude as the number of intervening performance periods increase (Arthur et al., 2007a; Henry & Hulin, 1987; Humphreys, 1960; Ryan & Connell, 1989). Although the correlations between most adjacent performance scores were low in magnitude, these correlations decreased as the temporal distance between two performance periods decreased.

Therefore, given the statistically significant and high adjacent session intercorrelations, it was expected to become more difficult for other variables like self–efficacy to explain subsequent performance over and above past performance, thus, providing support for Hypothesis 3.

Table 9

Hierarchical Regression Results Testing the Incremental Contribution of Self–Efficacy Beyond Previous Performance

Variable Added	**β**	R^2	ΔR^2
Criterion: Performance score Session 2			
Performance score Session 1	0.84	.01	
Self–Efficacy Time 1	0.95	.02	.01
Criterion: Performance score Session 5			
Performance score Session 4	2.60*	.07*	
Self–Efficacy Time 2	1.36*	.10**	.03**
Criterion: Performance score Session 10			
Performance score Session 9	4.91**	.13**	
Self–Efficacy Time 3	2.63*	.13**	.00
Criterion: Performance score Session 12			
Performance score Session 11	2.02*	.06*	
Self–Efficacy Time 4	0.72	.06*	.00

Note: $N = 198$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 10

Performance Session Intercorrelations across All Sessions

	Performance (Jane's Fleet Command Session Number)																
Session	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	—																
1	.12	—															
2	.12	.08	—														
3	.16*	.22**	.33**	—													
4	.05	.20**	.20**	.29**	—												
5	.01	.19**	.08	.26**	.27**	—											
6	.08	.18**	.12	.29**	.34**	.54**	—										
7	-.03	-.02	.06	.17*	.21**	.27**	.33**	—									
8	-.08	.06	.08	.19**	.24**	.35**	.35**	.55**	—								
9	-.09	.11	.16*	.05	.17*	.25**	.33**	.38**	.42**	—							
10	.02	.10	.06	.06	.18**	.22**	.26**	.28**	.38**	.38**	—						
11	.04	.06	.01	.05	.01	.06	.09	.10	.15*	.21**	.15*	—					
12	.07	.13	.16*	.24**	.16*	.25**	.25**	.19**	.25**	.27**	.28**	.38**	—				
13	.16*	.23**	.12	.19*	.26**	.28**	.20**	.31**	.31**	.35**	.40**	.52**	.48**	—			
14	-.16*	.13	.05	.15*	.20**	.25**	.21**	.26**	.26**	.35**	.39**	.39**	.37**	.40**	—		
15	.02	.12	.12	.16*	.17*	.20**	.21**	.24**	.28**	.30**	.32**	.32**	.21**	.35**	.36**	—	
16	-.02	.01	.09	.07	.01	.05	.04	-.01	-.02	.01	.01	.06	.06	.22**	.24**	.01	—

Note. Session 0 = baseline, $N = 198$. * $p < .05$. ** $p < .01$.

DISCUSSION

The primary objective of the current study was to investigate the relationship between training performance and self–efficacy using a longitudinal design (approximately 11 weeks) in the context of massed and distributed practice. This study was also conducted to address ambiguities in self–regulation theories (e.g., social–cognitive theory; Bandura, 1997) and other empirical literature (e.g., Ackerman et al., 1995; Arthur et al., 2007a; Heggestad & Kanfer, 2005; Judge et al., 2007; Vancouver & Kendall, 2006) regarding the performance and efficacy relationship and the role of performance and efficacy within the context of training. Consistent with the theories (Björk, 1994; Glenberg, 1979; Hull, 1943; Shea & Morgan, 1979; Wulf & Shea, 2002) and the meta–analyses (e.g., Donovan & Radosevich, 1999) presented in the introduction, the first set of hypotheses (Hypotheses 1a and 1b) predicted that the performance–efficacy relationship would vary as a function of the spacing of practice protocol (i.e., *distributed* and *massed*) such that the relationship in the distributed protocol during acquisition and reacquisition will be stronger than that in the massed protocol.

Overall, the results presented at the between– and within–person level of analysis using HLM failed to provide any support for these hypotheses. Furthermore, the hypothesis (Hypothesis 2) that the performance and efficacy relationship will be stronger when the practice condition (distributed or massed) during acquisition is the same as during reacquisition was also not supported. However, when past performance was considered, the unique contribution of self–efficacy to subsequent task performance decreased over time, thus providing support for the final hypothesis (Hypothesis 3). In the subsequent sections, theoretical and practical implications of these findings are discussed.

Theoretical Implications

Despite consistent findings in the performance and efficacy literature (e.g., Vancouver & Kendall, 2006; Vancouver et al., 2001; Vancouver et al., 2002) that posit a positive relationship between performance and efficacy at the between–persons level of analysis, the present study found weak and inconsistent relationship between performance and efficacy at the between–persons level of analysis in the context of spacing of practice protocols.

Furthermore, the discrepancy between the present findings and the literature on the positive correlational relationship between performance and efficacy may be attributed to the integration of spacing of practice protocols and the longitudinal design of the present study. Specifically, Vancouver and his colleagues (Vancouver et al., 2001; Vancouver et al., 2002; Vancouver & Kendall, 2006) did not include practice protocols (distributed and massed) and a nonpractice interval between acquisition and reacquisition while examining the relationship between performance and efficacy over time. Although practice protocols have received no attention in the performance and efficacy literature, adding these variables contributed further to the understanding of the performance and efficacy relationship over time. Specifically, integrating practice protocols in the performance–efficacy literature enhanced our understanding of the boundary conditions under which the performance and efficacy relationship evolves over time. The relationship between performance and efficacy did not vary as a function of practice protocols using a cognitively complex task. This finding has some support in the practice spacing literature, where scholars (Donovan & Radosevich, 1999) have found minimal differences in the effectiveness of one practice protocol over the other when using a cognitively complex task. Furthermore, explanations for this finding may also be found in the theoretical principles of Wulf and Shea's (2002) reconstruction hypothesis and Glenberg's (1979) component–process theory. These explanations are discussed in the subsequent pages.

When examining the performance and efficacy relationship at a within–person, across time level of analysis, Vancouver et al. (2001) found that self–efficacy was negatively related to subsequent performance. Furthermore, Vancouver and Kendall (2006) reported a significant negative relationship between self–efficacy and performance based on within–person analyses. Vancouver and his colleagues attributed their findings to the control theory perspective (Powers, 1973) and speculated that individuals' overconfidence in their abilities led to complacency which adversely affected their subsequent performance over time. Vancouver et al.'s findings were in contrast to Bandura's (1997) findings which revealed a significant positive relationship between performance and efficacy which is consistent with social cognitive theory. However, Bandura has predominantly used between–persons analyses to test his hypotheses. Although, a statistically significant relationship was not found between performance and efficacy in the present study, a weak and positive relationship emerged. Inconsistent with Vancouver et al.'s findings, these results suggest that the performance and efficacy relationship is still ambiguous at best when examined in the context of practice protocols and phases of learning at the within–person across time analyses.

One of the reasons for this ambiguity may be due to the longitudinal design of this study that consisted of one or two weeks of acquisition, an 8–week nonpractice interval, and one-half or one week of reacquisition. This longitudinal design is also an important contribution to the literature because it permitted an examination of the performance and efficacy relationship over a considerable long period of time. Unlike the longitudinal design used in the current study, researchers (e.g., Heggestad & Kanfer, 2005; Vancouver et al., 2001; 2002) in the past have used only one to two–hour long sessions with multiple trials to examine the performance and efficacy relationship using a within–person analysis. Furthermore, to my knowledge there is no study to date that has investigated the performance and efficacy relationship using an 8–week nonpractice interval.

Other explanations for these ambiguous findings, especially in the acquisition phase of learning may be attributed to the reconstruction hypothesis. Wulf and Shea (2002) posited that the contextual interference created by distributed schedules leads to forgetting of action plans during the initial trials of skill acquisition. The distributed practice schedules then require the repeated reconstruction of action plans—an activity that is less required in the massed schedules. One may speculate that as the participants in the current study were involved in other activities (e.g., starting new semester, managing their class schedule) between their initial trials of the distributed condition, they may not have had enough time to reconstruct their next action plan or strategy.

Because of the start of the school semester activities, it may also be argued that the time lag between the two trials was not long and optimal enough to provide distributed participants opportunity to think about their next action plan. Mitchell and James (2001) speculated that an optimal time lag between two variables is dependent on task complexity, the structure and spacing of practice presentations, and the rate of change between the two variables, in the current study—performance and efficacy. Wulf and Shea argued that because of already existing action plans in the working memory for participants in the massed schedule, massed schedules are more effective initially (e.g., during skill acquisition) than distributed schedules, whereas, distributed schedules are more effective during retention and transfer of skills. This might be another explanation for why the performance and efficacy relationship did not vary as a function of spacing of practice protocols in the acquisition phase of learning.

Based on the elaboration hypothesis (Shea & Morgan, 1979), it is also plausible that the participants may have considered massed schedules to be "distributed" and compared the new task information learned in the second hour of massed training with the task information in the first hour of massed training. This may have increased the level of distinctiveness which resulted in the massed schedule participants performing as good as distributed schedule participants. On the other hand, the participants in the distributed schedule may have failed to form associative processing between the two trials as the length between the two trials was too long. This is another probable explanation that may have accounted for the ambiguous results between the performance and efficacy relationship during the acquisition phase of learning.

Furthermore, to the best of my knowledge, the present study is one of the first studies that integrated spacing of practice and performance–efficacy literature to understand the rate of change between performance and efficacy longitudinally. Mitchell and James (2001) posited that one of the main conditions under which the theory directs methodology with respect to time is the time lag between two variables. In this study, the time lag for the distributed condition was larger compared to the time lag for massed. Furthermore, the time lag between self–efficacy administration 3 (performance at Session 9) and self–efficacy administration 4 (performance at Session 11) was an 8–week nonpractice interval. It could be argued here that this nonpractice interval which was an integral part of the study design mitigated the strength of the performance and efficacy relationship, thus, failing to provide support for the Hypotheses 1b. It may also be speculated here that the nonpractice interval was too long for the participants to sustain their attention on the previously learned task information. Hence, rather than reconstructing action plans and strategies for the upcoming sessions (based on the reconstruction hypothesis), participants may have forgotten significant task information or as Guthrie (1935) argued that learning did not disappear because of the length of time between two intervals, but primarily because of the *new* learning that occurred during that period of time. More research is needed to clarify these relationships over time, specifically in the context of nonuse periods. Furthermore, future research should continue to integrate practice protocols and phases of learning in the performance and efficacy relationship to enhance our understanding of performance and efficacy relationship over time.

Although the first two sets of hypotheses in the study were not supported, the results are consistent with other literature (Austin, 1921; Dempster, 1988; Donovan & Radosevich, 1999) that suggests that when the tasks have been more complex in nature, research has failed to show the superiority of distributed practice condition over massed. Relatedly, research investigating the effects of distributed practice on complex tasks has been limited and less conclusive (Goldstein & Ford, 2002). Hence, it is speculated here that cognitively complex task may be acting as a boundary condition in the effectiveness of distributed practice over massed, and therefore, failed to provide support for the key hypotheses.

However, the first set of hypotheses (Hypothesis 1a and 1b) are consistent with the meta–analytic study conducted by Donovan and Radosevich (1999). Specifically, Donovan and Radosevich suggested that practice spacing researchers should pay special attention to the importance of boundary conditions while considering the effectiveness of distributed protocols over massed. Donovan and Radosevich stated that very little is known about the complex cognitive tasks in the practice spacing literature. In their meta–analytic review, Donovan and Radosevich further posited that the type of task, the length of intertrial time interval, and the interaction between these two factors play a significant role in determining the magnitude of the spacing of practice effects. Specifically, Donovan and Radosevich found that as the intertrial interval between the distributed practice trials became shorter, the standardized mean differences between distributed and massed groups increased. They attributed this surprising result to the fact that any additional time between the distributed practice trials could have been harmful to the subsequent performance. In a similar vein, a plausible explanation for the finding—that the performance and efficacy relationship did not vary as a function of practice protocols—may be that the intertrial time intervals for the distributed protocols were too long to detect any interaction effects between the performance–efficacy relationship and practice protocol. Specifically, trainees in the distributed practice condition experienced eight breaks in training during the skill acquisition phase (seven 23-hour breaks).

All these explanations may have accounted for the lack of support found for the first set of hypotheses. In addressing the importance of time, George and Jones (2000) state that time should be treated as an essential element in theory building rather than boundary condition because “time is an intrinsic property of consciousness” (p. 659). George and Jones further explain that the stream of consciousness is ordered temporally and all conscious processing takes place within the flow of time. The longitudinal design of this study was an important contribution to the performance and efficacy literature and more research is needed to understand the complexity of the performance–efficacy relationship over time.

In conjunction with the theoretical principles of Glenberg’s (1979) component–process theory, it was also predicted that the performance and efficacy relationship will vary as a function of whether the practice condition during acquisition is the same or different from the practice condition during reacquisition such that the relationships are stronger when the practice condition is the same instead of when it is different. Specifically, it was predicted here that greater spacing of presentations (i.e., distributed schedules) combined with similar contextual components during acquisition and reacquisition will result in better encoding and hence, in stronger performance and efficacy relationships over time (Glenberg, 1979). This hypothesis was also examined at the within–person analysis using HLM. The results failed to show an interaction effect between performance and practice category ($\gamma = .01$, *ns*; see Table 7). These results were inconsistent with Glenberg's (1979) component–process theory.

One may speculate that the context during which the task information was presented in the beginning of the study (during skill acquisition) and before the 8–week nonpractice interval may not be similar to the context during which task information was presented after the 8–week nonpractice interval (during skill reacquisition). This is possible as data were collected in the beginning of the semester when students had just started classes (*before* 8–week nonpractice interval) and at the end of the semester when students were taking final exams (*after* 8–week nonpractice interval). Maybe this dissimilarity between the two contextual components before and after the nonpractice interval led to poor encoding of task information and hence, mitigated the performance and efficacy relationship after the nonpractice interval. Another explanation for this may be that other components (i.e., structural and descriptive) that Glenberg (1979) identified confounded the performance and efficacy relationship. It is also possible that the structural component representing relations and associations between tasks was affected by the time lag between intertrial intervals and the 8–week long nonpractice interval such that the fluctuations between the tasks were too large to result in greater encoding. Also, effective encoding of task information is dependent upon the degree to which the descriptive component (i.e., representing specific task features) of component–process theory is included in the memory trace (Glenberg, 1979). Because the task was a complex nonmotor task, it is also possible that the complexity of the task may have affected the strength of the performance–efficacy relationship.

Finally, the hypothesis that when past performance is controlled the incremental contribution of self–efficacy to subsequent task performance will decrease over time was strongly supported. The hierarchical regression results indicated that the self–efficacy ratings explained variance beyond prior performance and the change in R^2 was statistically significant only at Time 2. This finding is consistent with Heggestad and Kanfer's (2005) conclusions that when past performance is controlled, the unique effect of self–efficacy on task performance is substantially attenuated. This result is also consistent with Judge et al.'s (2007) meta–analytic findings that across all studies and moderator conditions, the incremental validity of self–efficacy on task and job performance was mitigated in the presence of specified individual difference variables (i.e., personality, general mental ability, experience). Furthermore, Arthur et al. (2007a) found that although the relationship between team efficacy and team performance improved over time, the team efficacy ratings failed to explain variance above previous team performance. Bandura (1997) defined team efficacy as a shared belief in the team's conjoint capabilities to organize and execute the courses of action required to produce given levels of attainments. Therefore, the finding is consistent with the position taken by many researchers (e.g., Ackerman et al., 1995; Arthur et al., 2007a; Heggestad & Kanfer, 2005; Judge et al., 2007) and that the positive relationship found between self–

efficacy and performance is due more to the effect of past performance on self–efficacy rather than the effect of self–efficacy on subsequent performance as posited by other scholars (e.g., Bandura, 1997; Bandura & Wood, 1989; Stajkovic & Luthans, 1998).

Consistent with Powers' (1973) control theory and Guthrie's (1935) recency principle, the finding suggests that self–efficacy will have very little incremental value above previous performance and will fail to predict subsequent performance over time. Based on the cybernetics structure of self–regulation, Powers argued that those who believe that they can meet their goals are *less* likely to work harder in meeting their goals as compared to those who do not believe in their ability to meet their goals. Specifically, Powers posited that over time, self–efficacy will fail to explain incremental variance above previous performance. Powers' control theory is in stark contrast to Bandura's (1997) social cognitive theory. Furthermore, with reference to the recency principle, Guthrie (1935) posited that when individuals are confronted with a stimulating situation or experience that closely resembles an earlier one, the individuals are more likely to react as they did previously. This argument is reflected in the simplex–like pattern that suggests immediate past performance should be more closely related to current performance as the training progresses. According to simplex–like pattern, the correlations between adjacent trials are strongest and they become smaller the further apart the trials are in the sequence (Humphreys, 1960). Therefore, immediate past performance should be more predictive of current performance as training progresses. Guthrie's (1935) recency principle also complements Bandura's (1997) enactive mastery that posits individuals will use their recent successful experiences as compared to distal experiences to formulate their efficacy beliefs. Hence, the findings from this hypothesis are consistent with the effects posited by Guthrie's recency principle, Power's control theory, and the simplex–like pattern.

Practical Implications

Although two of the three hypotheses were not supported, the findings have some important practical implications in the real-world training contexts. Specifically, in designing a training program, the training practitioners should be aware of the boundary conditions associated with the effectiveness of distributed practice over massed. One of the boundary conditions that the trainers should be especially aware of is the time lag between two trials and variables. Many scholars (Ancona et al., 2001; Donovan & Radosevich, 1999; Mitchell & James, 2001) have addressed the importance of time lag by suggesting that both time lag between intertrial intervals and that between two variables is critical to our understanding of the relationships between variables over time. For example, the trainers may be able to strengthen or weaken the relationship between performance and efficacy by manipulating the time lag between two trials. This implication is important in the real world especially when time and resources are not easily accessible for training purposes. Furthermore, trainers should be very cautious in accepting the effectiveness of distributed protocol over massed as a universal doctrine. Relatedly, Donovan and Radosevich (1999) state that the superiority of distributed over massed protocols "is not as strong or pervasive as many researchers in the past have been inclined to accept" (p. 802). More research is needed to investigate the performance and efficacy relationship longitudinally and within the context of different practice protocols.

On a related note, trainers should also be aware of the duration of nonpractice period or the period when skills are minimally used to examine the rate of change between performance and efficacy of the trainees. For example, there are situations where individuals receive initial training on skills and knowledge that they may not be required to use or may not have the opportunity to perform for extended periods of time. Reserve personnel in the military may receive formal training only once or twice a year with the expectation that they will only need a limited amount of refresher training to reacquire any skill that has been lost when they are called up for active duty (Arthur et al., 2007b; Wisher, Sabol, Hillel, & Kern, 1991). In a similar vein, there are other examples of disaster and rescue teams whose skills are not utilized until the time of emergency or a natural catastrophe like a tsunami. A thorough understanding of the nonpractice interval and performance–efficacy relationship in the context of practice spacing may enhance the post-training skill retention. This study provides a preliminary outlook in understanding these relationships using a cognitively complex skill and a longitudinal design.

In the education industry, there is a growing demand for trainers to train non–native English population in English language proficiency skills (Minaya-Rowe, 2004; Wyra, Lawson, & Hungi, 2007). A sound understanding of the performance and efficacy relationship and spacing of practice presentations will provide substantial contribution to the education industry. Furthermore, apart from the language proficiency training,

training in math and reading comprehension may also be provided to school children by using the principles of practice spacing and the self–regulation (Bandura, 1997) and control (Powers, 1973) theories.

Future Research

In addition to the future research briefly discussed thus far, one potentially fruitful area of investigation is the interaction between individual difference variables and performance and efficacy relationship in the context of practice spacing. Judge et al. (2007) showed that across all studies and moderator conditions, the incremental validity of self–efficacy on task and job performance was attenuated in the presence of individual difference variables (i.e., personality, general mental ability, experience). Many research questions can be developed to investigate if these meta–analytic findings can be generalized when examined longitudinally and in the context of practice spacing. For example, one potential research question that can be asked is how do motivational variables interact with performance and efficacy in the context of practice spacing? Furthermore, what effect do individual difference variables (e.g., conscientiousness, general mental ability) have on the performance and efficacy relationship when examined after long nonuse periods?

Relatedly, one may examine if certain motivational variables like learning and performance goal orientation moderate the relationship between performance and efficacy over a long nonuse period. It is also fruitful to investigate if self–efficacy contains motivational properties not reflected in prior performance that have more influence on performance after extensive nonuse rather than more immediate performance. These research questions may be examined by increasing the duration of skill acquisition and reacquisition to examine the interaction of different individual difference variables and self–efficacy over a period of 12 to 24 months with longer nonuse periods. In doing so, it is possible to increase the self–efficacy administrations at level one to examine the within–person effects using HLM.

Moreover, future research should also attempt to directly manipulate self–efficacy in a training context to assess the impact of self–efficacy intervention on subsequent performance and other self–regulatory processes over time. Furthermore, the effect of spacing of practice protocol on the performance and efficacy relationship should also be explored in field settings (e.g., firefighters, rescue teams), in order to determine the generalizability of the current set of results generated in a laboratory setting. As mentioned earlier in practical implications section, it is important to continue investigating the performance and efficacy relationship in the acquisition of language skills in the education industry. This is an important area of future research because in recent years there has been an increasing demand to train non–native English speakers in the acquisition of English language skills in all grades of the school system as well as college (O’Malley, Ayala, Zilbert, & Cernohous, 2007). Moreover, the federal and state governments are funding and promoting research in test development and assessment of Texas English Language Proficiency Assessment System (TELPAS). An empirical investigation of the performance and efficacy relationship in the context of language skill acquisition will further our understanding of the performance–efficacy relationship in a different field setting. This is another important avenue for future research.

Another fruitful area of future research is the influence of practice spacing (both distributed and massed) on the team performance and team efficacy relationship. Arthur et al. (2007a) investigated the relationships between team efficacy operationalizations (additive and referent-shift consensus operationalizations) and team performance over a 2-week period. Additive operationalization includes aggregating team members' ratings of their individual capabilities for a particular task, whereas, referent-shift consensus operationalization includes aggregating team members' appraisals of their team's capability within a specified domain (Arthur et al., 2007a). Arthur et al. found that although both operationalizations were related to team performance, additional analyses provided support for the superiority of the referent-shift operationalization. Furthermore, the reference-shift consensus operationalization explained additional variance in team performance beyond the additive operationalization but the converse was not true. Hence, future research may investigate the effect of spacing of practice protocols (distributed and massed) and the two phases of learning (skill acquisition and reacquisition) on the relationship between these team efficacy operationalizations and team performance over time. Thus, it seems worthwhile to investigate whether the boundary conditions applied to the performance and efficacy relationship at the individual level generalize to the team level.

Limitations

One of the limitations of the present study arises from its methodology. The self–efficacy measure was not administered immediately following the 8–week nonpractice interval before performance Session 10. However, self–efficacy ratings were collected immediately following performance Session 11 after nonpractice interval. Because of this limitation in the design, it was not possible to examine the rate of change in the efficacy scores when the participants returned to the lab after eight weeks of nonuse period. Specifically, it was not possible to examine whether the efficacy ratings increased, decreased, or remained stable after the nonuse period. Relatedly, the current study relied on natural variation within individuals over time to assess the efficacy effects. Given the relatively few observations of self–efficacy (three for acquisition and two for reacquisition) per person in each phase of learning, little variance was available to affect the performance and efficacy relationship over time. Perhaps, directly manipulating self–efficacy would better test the first two sets of hypotheses.

On a related note, another limitation of the current study pertains to the number of self–efficacy measure administrations (i.e., five) at the level one of the within–person HLM analysis. Vancouver and Kendall (2006) used a similar number of efficacy measurements and found significant negative effects of self–efficacy on performance. However, they did not include the two phases of learning (i.e., skill acquisition and reacquisition) in their study. After incorporating these two phases of learning in the present study, self–efficacy measure administrations were reduced to three during skill acquisition and two during skill reacquisition. Maybe future research should consider integrating more administrations of a self–efficacy measure at level one of HLM analysis.

Conclusion

Past research on the effectiveness of distributed protocols over massed when using cognitively complex tasks is not conclusive. Furthermore, the joint effects of performance and efficacy relationship in the training context over time have also resulted in mixed conclusions (Bandura, 1997; Vancouver & Kendall, 2006). This study sought to contribute to both, the practice spacing and performance and efficacy literature by clarifying some of the past conflicting conclusions. Specifically, the present study sought to investigate the relationship between training performance and self–efficacy using a longitudinal design in the context of massed and distributed practice. The results did not provide support for the first set of hypotheses that the performance–efficacy relationship will vary as a function of the spacing of practice protocol (i.e., *distributed* and *massed*). The relationship in the distributed protocol during acquisition and reacquisition was not stronger than that in the massed protocol.

Furthermore, inconsistent with Glenberg's (1979) component–process theory, the performance and efficacy relationship did not vary as a function of whether the practice condition during acquisition was the same or different from the practice condition during reacquisition. On the other hand, and consistent with the Power's (1973) control theory and Guthrie's (1935) recency principle, when past performance was controlled self–efficacy failed to explain additional variance in subsequent task performance beyond past performance. This finding is consistent with the research (Ackerman et al., 1995; Arthur et al., 2007a; Shea & Howell, 2000) that has concluded that the best predictor of current performance is immediate past performance and that self–efficacy fails to provide any incremental value to the current performance above past performance. The findings from the current study provide a stepping stone to start integrating spacing of practice protocols in the performance and efficacy literature and provide new avenues for performance and efficacy longitudinal research.

REFERENCES

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied, 1,* 270-303.

Ancona, D. G., Okhuysen, G. A., & Perlow, L. A. (2001). Taking time to integrate temporal research. *Academy of Management Review, 26,* 512-529.

Arthur, W. Jr., Bell, S. T., & Edwards, B. D. (2007a). A longitudinal examination of the comparative criterion–related validity of additive and referent–shift consensus operationalizations of team efficacy. *Organizational Research Methods, 10,* 35-58.

Arthur, W. Jr., & Bennett, W. Jr. (1996). *Skill retention and decay: A meta-analysis [Final report]*. (AL/HR-TR-1996-XXXX). AL/HRTD contract F41622-93-M-2342, Technical Training Division, Brooks AFB, TX.

Arthur, W. Jr., Day, E. A., Bennett, W. Jr., McNelly, T. L., & Jordan, J. A. (1997). Dyadic versus individual training protocols: Loss and reacquisition of a complex skill. *Journal of Applied Psychology, 82,* 783-791.

Arthur, W. Jr., Day, E. A., Villado, A. J., Boatman, P. R., Bhupatkar, A., Kowollik, V., & Bennett, W. Jr. (2007b, April). *Decay and reacquisition of a complex skill: An investigation of practice schedules, observational rehearsal, and individual differences.* In F. A. Oswald and E. M. Oberlander (Co–Chairs), Adaptive Skills and Adaptive Performance: Today's Organizational Reality. Symposium presented at the 22nd Annual Conference of the Society for Industrial and Organizational Psychology, New York City, NY.

Arthur, W. Jr., Villado, A. J., Boatman, P. R., Bhupatkar, A., Day, E. A., & Bennett, W. Jr. (2005, April). *Complex skill acquisition, decay, and transfer: The comparative effectiveness of massed and distributed practice schedules.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Arthur, W. Jr., Villado, A. J., Boatman, P. R., Bhupatkar, A., Day, E. A., Kowollik, V., & Bennett, W. Jr. (2006). *Standards for readiness assessment and tracking of warfighter training systems research.* The Defense Advanced Research Project Agency (DARPA), and Air Force Research Laboratory Human Effectiveness Directorate (AFRL/HE), Mesa, AZ. Contract No. F41624–97–D–5000, L–3 Communications Corporation, Link Simulation and Training Division, Mesa, AZ.

Austin, S. D. M. (1921). A study in logical memory. *American Journal of Psychology, 32,* 370-403.

Baddeley, A. D. (1986). *Working memory.* New York: Oxford University Press.

Bandura, A. (1977). Self–efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215.

Bandura, A. (1982). Self–efficacy mechanism in human agency. *American Psychologist, 37,* 122-147.

Bandura, A. (1986). *Social foundations of thought and action: A social–cognitive view.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self–efficacy: The exercise of control.* New York: Freeman.

Bandura, A., & Locke, E. A. (2003). Negative self–efficacy and goal effects revisited. *Journal of Applied Psychology, 88,* 87-99.

Bandura, A., & Wood, R. (1989). Effect of perceived controllability and performance standards on self–regulation of complex decision making. *Journal of Personality and Social Psychology, 56,* 805-814.

Battig, W. F. (1966). Facilitation and interference. In E. A. Bilodeau (Ed.), *Acquisition of skill* (pp. 215-244). New York: Academic Press.

Björk, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognitions: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Björk, R. A., & Björk, E. L. (1992). A new theory of disuse and an old theory of stimulation fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (pp. 35-67). Hillsdale, NJ: Erlbaum.

Borkovec, T. D. (1978). Self–efficacy: Cause or reflection of behavioral change. *Advances in Behavioral Research and Therapy, 1,* 163-170.

Bower, G. H. (1972). Stimulus–sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 221-246). Washington, DC: Wiley.

Bretz, R. D., & Thompsett, R. E. (1992). Comparing traditional and integrative learning methods in organizational training programs. *Journal of Applied Psychology, 77,* 941-951.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Corrigan, P. W. (1990). Theoretical weakness in behavior theory is no more than statistical variance: Response to Christina Lee. *Journal of Behavior Therapy and Experimental Psychiatry, 21,* 141-142.

Daft, R. L., & Lengel, R. H. (1984). Information richness: A new approach to managerial behavior and organizational design. *Research in Organizational Behavior, 6,* 191-233.

DeCecco, J. P. (1968). *The psychology of learning and instruction: Educational psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43,* 627-634.

Donovan, J. J., & Radosevich, D. J. (1999). A meta–analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84,* 795-805.

Earley, P. C., & Lituchi, T. R. (1991). Delineating goal and efficacy effects: A test of three models. *Journal of Applied Psychology, 76,* 81-98.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology.* Teachers College, Columbia University, NY: Columbia University Press.

Ericsson, K. A., & Charness, N. (1994). Expert performance: It's structure and acquisition. *American Psychologist, 49,* 725-747.

Ericsson, K. A., Krampe, R. T., & Tesch–Roemer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363-406.

George, J. M., & Jones, G. R. (2000). The role of time in theory and theory building. *Journal of Management, 26,* 657-684.

Glenberg, A. M. (1979). Component–levels theory of the effects of spacing of repetitions on recall and recognition, *Memory & Cognition, 7,* 95-112.

Goldstein, I. M., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Belmont, CA: Wadsworth/Thompson Learning.

Guthrie, E. R. (1935). *The psychology of learning.* New York: Harper & Row.

Hawkins, R. M. (1992). Self–efficacy: A predictor but not a cause of behavior. *Journal of Behavior Therapy and Experimental Psychiatry, 23,* 251-256.

Heggestad, E. D., & Kanfer, R. (2005). The predictive validity of self–efficacy in training performance: Little more than past performance. *Journal of Experimental Psychology: Applied, 11,* 84-97.

Henry, R. A., & Hulin, C. L. (1987). Stability of skilled performance across time: Some generalizations and limitations on utilities. *Journal of Applied Psychology, 72,* 457-462.

Hertenstein, E. J. (2001). Goal orientation and practice conditions as predictors of training results. *Human Resource Development Quarterly, 12,* 403-409.

Hull, C. L. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

Humphreys, L. G. (1960). Investigations of simplex. *Psychometrika, 25,* 313-323.

Immink, M. A., & Wright, D. L. (1998). Contextual interference: A response planning account. *The Quarterly Journal of Experimental Psychology, 51A,* 735-754.

James, W. (1890). *Principles of psychology.* New York: Holt.

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B., & Rich, B. L. (2007). Self–efficacy and work–related performance: The integral role of individual differences. *Journal of Applied Psychology, 92,* 107-127.

Krug, D., Davis, T. B., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology, 82,* 366-371.

Lee, C. (1989). Theoretical weakness lead to practical problems: The example of self–efficacy theory. *Journal of Behavior Therapy and Experimental Psychiatry, 20,* 115-123.

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects re–considered. *Research Quarterly, 59,* 277-287.

Lindsley, D. H., Brass, D. J., & Thomas, J. B. (1995). Efficacy–performance spirals: A multilevel perspective. *Academy of Management Review, 20,* 645-678.

Locke, E. A., & Latham, G. P. (1994). Goal setting theory. In H. F. O'Neil and M. Drillings (Eds.), *Motivation: Theory and research* (pp. 13-29). UK: Lawrence Earlbaum.

Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science, 9,* 221-239.

McGeoch, G. O. (1935). The conditions of reminiscence. *American Journal of Psychology, 47,* 65-89.

Minaya-Rowe, L. (2004). Training teachers of english language learners using their students' first language. *Journal of Latinos and Education, 3,* 3-24.

Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review, 26,* 530-547.

O'Malley, K. J., Ayala, L., Zilbert, E., & Cernohous, J. (June, 2007). *Assessments for English language learners: Validity evidence from California and Texas.* Paper presented at the Council of Chief State School Officers national meeting, Nashville, TN.

Powers, W. T. (1973). *Behavior: The control of perception.* Chicago: Aldine.

Richard, E. M., Diefendorff, J. M., & Martin, J. H. (2006). Revisiting the within-person self-efficacy and performance relation. *Human Performance, 19,* 67-87.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internationalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57,* 749-761.

Sadri, G., & Robertson, I. T. (1993). Self–efficacy and work–related behavior: A review and meta–analysis. *Applied Psychology: An International Review, 42,* 139-152.

Schmidt, R. A., & Björk, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts in training. *Psychological Science, 3,* 207-217.

Schneider, W. (1985). Training high–performance skills: Fallacies and guidelines. *Human Factors, 27,* 285-300.

Shea, C. M., & Howell, J. M. (2000). Efficacy–performance spirals: An empirical test. *Journal of Management, 26,* 791-812.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 179-187.

Shea, J. B., & Zimny, S. T. (1983). Contextual effects in memory and learning movement information. In R. A. Magill (Ed.). *Memory and control of action* (pp. 145-166). Amsterdam: North-Holland.

Shebilske, W. L., Goettl, B. P., Corrington, K., & Day, E. A. (1999). Interlesson spacing and task-related processing during complex skill acquisition. *Journal of Experimental Psychology: Applied, 5,* 413-437.

Sonalysts, Inc. (2000). *Jane's Fleet Command* [Computer software]. Redwood City, CA: Electronic Arts.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 3,* 641-656.

Stajkovic, A. D., & Luthans, F. (1998). Self–efficacy and work–related performance: A meta–analysis. *Psychological Bulletin, 124,* 240-261.

Vancouver, J. B., & Kendall, L. N. (2006). When self–efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology, 91,* 1146-1153.

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self–efficacy on performance. *Journal of Applied Psychology, 87,* 506-516.

Vancouver, J. B., Thompson, C. M., & Williams, A. A. (2001). The changing signs in the relationships among self–efficacy, personal goals, and performance. *Journal of Applied Psychology, 86,* 605-620.

Wexley, K. M., & Latham, G. P. (2002). *Developing and training human resources in organizations.* Upper Saddle River, NJ: Prentice Hall.

Whetzel, J. H., Arthur, W. Jr., & Volz, R. A. (in press). The effectiveness and efficacy of intelligent agents as team training partners in the acquisition of complex skills in a game–based training environment. In H. F. O'Neil, & R. Perez (Eds.), *Computer games and adult learning.* The Netherlands: Elsevier.

Wisher, R. A., Sabol, M. A., Hillel, H., & Kern, R. P. (1991). *Individual ready reserve (IRR) call-up: Skill decay*. (ARI Research Rep. No. 1595). Alexandria, VA: U.S. Army Research Institute.

Wood, R. E., & Bandura, A. (1989). Impact of conceptions of ability on self–regulatory mechanisms and complex decision making. *Journal of Personality and Social Psychology, 56,* 407-415.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review, 9,* 185-211.

Wyra, M., Lawson, M. J., & Hungi, N. (2007). The mnemonic keyword method: The effects of bidirectional retrieval training and of ability to image on foreign language vocabulary recall. *Learning and Instruction, 17,* 360-371.

APPENDIX A

Self–Efficacy Scale

Please read each of the statements listed below and indicate how much you personally agree with each statement by marking the response that most applies to you.

Name: ____ Date: ________

①	②	③	④	⑤
Strongly Disagree	**Disagree**	**Neither Agree nor Disagree**	**Agree**	**Strongly Agree**

1. I feel confident in my ability to perform well on Fleet Command.	① ② ③ ④ ⑤
2. I can meet the challenges of Fleet Command.	① ② ③ ④ ⑤
3. I know I can achieve good scores at Fleet Command.	① ② ③ ④ ⑤
4. I know that I can master Fleet Command.	① ② ③ ④ ⑤
5. I do not think Fleet Command is something that I will become good at.	① ② ③ ④ ⑤
6. I am confident that I have what it takes to perform Fleet Command well.	① ② ③ ④ ⑤
7. I know that I am capable of getting better at Fleet Command.	① ② ③ ④ ⑤
8. I am confident that Fleet Command will seem less challenging to me when I have completed this study.	① ② ③ ④ ⑤
9. I am certain that I could cope with Fleet Command if it became more complex.	① ② ③ ④ ⑤
10. I know I could handle Fleet Command if it became more difficult.	① ② ③ ④ ⑤
11. I know I could succeed at Fleet Command if aspects of the game were altered.	① ② ③ ④ ⑤
12. If Fleet Command got any harder, I think it would be impossible for me to get a good score.	① ② ③ ④ ⑤

VITA

Name: Alok Bhupatkar

Address: Pearson, Psychometric Services
Research Scientist
400 Center Ridge Drive, Suite F
Austin, TX 78753

Email Address: alok.bhupatkar@pearson.com

Education

Ph.D. Texas A&M University, College Station, Texas, 2007
Major: Industrial/Organizational Psychology

M.S. Emporia State University, Emporia, Kansas, 2003
Major: Industrial/Organizational Psychology

B.A. University of Poona, Poona, India, 1999
Major: Psychology

Professional Experience

January 2007–August 2007
<u>Intern</u>–Kenexa Corporation

- Conduct validation studies, analyze validation data, and examine the possibility for adverse impact; assemble and statistically analyze selection assessment data from clients.

Selected Awards

Fall 2004
Best graduate research poster award at the 1st Year Graduate poster competition at the Department of Psychology, Texas A&M University.

Spring 2003
Outstanding Graduate Student in Industrial Psychology Award at Teacher's College, Emporia State University.

Fall 2002
First recipient of the Harry Levinson scholarship in Organizational Behavior, Emporia State University.